Social Influence
Identification, Effect and Extensions

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Radu Petru Tănase
from Romania

approved in April 2018 at the request of

Prof. Dr. René Algesheimer
Prof. Dr. Claudio J. Tessone
Prof. Dr. Anne Scherer

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 04.04.2018
The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

To my family and my friends . . .

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Understanding how people influence or are influenced by their peers can help us understand the flow of market trends, product adoption and diffusion processes. The tremendous impact of social influence mechanisms can been seen in recent events like political elections in the US and UK, anti-vaccination campaigns in Eastern Europe and countless product promotions through influencer marketing campaigns. In 2016 influencer marketing was rated the fastest-growing customer acquisition channel [2], with 84% of marketers planning to launch at least one influencer marketing campaign in the following 12 months [1]. Still in 2016, nearly 40% of Twitter users made at least one purchase based on a tweet from an influencer [4], 70% of teenage users on Youtube said they are influenced more by "YouTubers" than by traditional celebrities [5] and 70% of millennials stated their purchase decisions were influenced by their peers [3].

Given the tremendous impact, there is extensive literature investigating social influence processes and the emergence of influencers. Starting with the seminal work of Katz and Lazarsfeld [73] on the two step model of communication flow between the mass media and the public, an important stream of research has been devoted to identifying a minority of individuals (called opinion leaders, influentials or influencers), that can influence the opinions and behavior of a large number of people. To identify influencers, early works relied on conducting surveys. People were asked either to state how influential they perceive themselves [e.g. 32] or to name who had influenced them in making a decision [e.g. 73]. The influencers were then identified as either the individuals with the highest self reported opinion leadership measure or by their position in the derived influence network. Facilitated by high data availability, modern approaches aim to infer how influential an individual is from traces left in the digital space. Nowadays researchers and practitioners typically assume a set of features believed to characterize influencers and then identify the influencers as the individuals with the highest values of these features. Such features range from psychological traits [139] to expertise [84, 109] or position in the social network (e.g. betweenness centrality [43], eigenvector centrality [21], node accessibility [130], k-shell [79], dynamical influence [80], expected force [83] or collective influence [98]). However, an important challenge of this approach is that there is hardly any consensus on which is the best set of features which describe influencers, with new studies continuously challenging previous findings. Furthermore, most methods have been developed for time-aggregated datasets, which neglects the inherent dynamic nature of influence relationships.

In the first study we contribute to this line of work by showing that influencer identification can be simplified if mapped to a wisdom of the crowds problem. We construct a framework in which individuals in a social group repeatedly evaluate the contribution of other members according to what they perceive as valuable. This approach does not require an apriori specification of the set of features describing the influencers, but lets each individual decide on his own, based on the preferences and beliefs held at that point in time. To aggregate the individual evaluations into a collective judgment, we develop an aggregation method that: (1) takes advantage of the temporality of the data; (2) addresses different sources of heterogeneity specific to social systems and (3) leads to results that are easily interpretable and comparable within and across systems. To illustrate our approach, we collected a large dataset of more than two million users contributing more than fifty million posts to online news discussions from three large, independent news providers. We show that following the approach we propose, it is straightforward to reveal those users who are consistently the most influential. Furthermore, we find that most influencers are influential within only one topic, which implies that influence is context dependent. This if often neglected in extant literature, where the identification of influencers is done based on one, often structural, feature.

In addition to the identification of influencers, an important stream of research was devoted to understanding the mechanisms through which individuals influence others and to the quantification of this effects. Seminal studies include Katz and Lazarsfeld [73] which show that the impact of social influence on brand switching is greater than that of traditional media, Coleman [30] and later Iyengar et al. [65] which show there is a positive effect of social contagion on new product adoption or Kim et al. [77] which show that social influence can accelerate the implementation of community health programs. The amount of evidence towards social contagion has led Godes [49] to argue for a shift of research focus from showing that social influence exists to understanding how it operates. We contribute to this line of research by investigating how social influence interacts with the individual tendency to engage in varied behavior. In the second study we construct a theoretical framework which relates the probability to repurchase a product to variety seeking behavior and social exposure. Within our framework, a purchase decision is represented as an unobserved two stage decision process. In the first stage a decision to repurchase or switch to a new product is made based on intrapersonal stimuli (variety seeking). Then, in the second stage, this decision is confronted with the social group (interpersonal stimuli, social exposure). To illustrate the approach we conducted two randomized controlled trials and an empirical study. Our results show that variety seeking has a negative effect on the probability to repurchase, social exposure having a positive valence towards repurchase has a positive effect while social exposure having a positive valence towards product switching has a negative effect. We illustrate the implications of our theoretical framework by constructing a Markov model of product choice in which we relate the probability to buy a product, either as a repurchase or new purchase, to variety seeking and social exposure. The results show that when both variety seeking and social exposure have a positive valence towards product switch, the effect of social exposure on product switch is overestimated as the decision to switch is already made by the consumer and thus social exposure has merely a reinforcement effect. On the other hand, when variety seeking has a positive valence towards product switch and social exposure has a positive valence towards repurchase, a repurchase reflects an actual influence effect, as in the absence of social exposure the individual would have switched to the new product. This is to our knowledge the first study documenting such an effect and shows that individuals who do not promote opinion or behavioral change can be as important for diffusion processes as are those who promote it.

In the third study we investigate how tools for influencer identification in social networks can be extended to understand and eventually solve public policy problems. Examples of public policy problems include preventing urban decay [39], improving health care delivery for patients with chronic illnesses [60] or improving effectiveness of highway construction projects [61]. Typically a public policy problem can be modeled as a set of variables (e.g. population, mobility, etc.) linked by causal relations between them. Once a problem is modeled, simulation scenarios are developed in order to understand the effect of certain policies (changes to variables in the model) on variables of interest. However, especially in large systems, an important and still open problem is the identification of model levers, that is, variables that can most efficiently and effectively control the entire system. We develop a framework in which a policy problem is mapped to a directed network, where nodes represent variables and links the causal relationships between them. Then, building on network controllability theory, we identify the model levers as the nodes that are the most influential in controlling the entire network. To illustrate our approach we analyze a classic system dynamics model: the World Dynamics model [40]. We show that changing the most influential levers has a significant effect on the variables of interest, while changing the least influential variables does not result in any change. A summary of the three studies together with their contribution can be found in Table 1.1.

The thesis is structured as follows. In Chapter 2 we review extant work on social influence. In Chapter 3 we present our approach to the identification of influencers. Then in Chapter 4 we discuss the theoretical framework relating repurchase behavior to variety seeking and social exposure. In Chapter 5 we present how tools for influencer identification in social networks can be used to solve public policy problems. Chapter 6 concludes the thesis with a discussion of the three studies and avenues for further research. Supporting material accompanying all chapters can be found in the appendix.

Study 1: The identification of influencers through the wisdom of the crowds

Research questions	• Who are the influential individuals in social interactions?
Core contributions	• Novel method to identify influencers. • Novel method to aggregate the individual evaluations into a collective judgment that considers the temporal variation of influence processes. • The identification method does not rely on predefined features. • Scalable aggregation method that can be applied to any temporal system where a performance metric is measured over time.
Data basis	• News discussion on three large, independent providers: CNN (2012 - 2014); The Atlantic (2007 - 2016); The Telegraph (2006 - 2016).
Methods	• Own implementation.
Main results	• Members of the social group agree on who are the influential individuals. • Influence is context dependent.

Study 2: The effects of social influence and variety seeking on repurchase behavior

Research questions	• Which is the differential impact of social influence supporting product repurchase (hidden influence) and social influence supporting product change (visible influence) on the probability to buy the advocated product?
Core contributions	• New conceptualization of social influence. • First study to show the positive effect of social influence on repurchase behavior in consumption situations characterized by variety seeking behavior. • In such situations, the total social influence effect reported in the literature might be under/overestimated.
Data basis	• Experimental: Two Amazon Mechanical Turk studies involving 464/1304 participants. • Empirical: User interactions in an online community where members share their cooking experience (2011 - 2015, 1.4 Mil. users).
Methods	• Experimental design; Mixed effects logistic regression; Markov model.
Main results	• Hidden influence has a positive effect on the probability to repurchase. • Visible influence has a negative effect on the probability to repurchase. • When both variety seeking and social influence have a positive valence towards product switch, the effect of social influence on product switch might be overestimated as the decision to switch is already made by the consumer and thus social influence has merely a reinforcement effect. • When variety seeking has a positive valence towards product switch and social influence has a positive valence towards repurchase, a repurchase reflects an actual influence effect, as in the absence of social influence the individual would have switched to the new product.

Study 3: Controlling complex policy problems

Research questions	• How to identify model levers (i.e. variables that can most efficiently and effectively control the entire system) in large dynamical systems?
Core contributions	• Method to identify model levers in large dynamical systems. • Combine system dynamics with network controllability to facilitate policy design.
Data basis	• World dynamics model [38].
Methods	• Network controllability analysis; Control centrality; Simulations; System dynamics models.
Main results	• It is enough to control only 53% of the variables to steer the entire system to an arbitrary final state. • Changing the most influential levers leads to a significantly larger change in the variables of interest compared to changing the least influential ones.

Table 1.1 Studies overview

Chapter 2

Background social influence

Social influence can be defined as "any process whereby a person's attitudes, opinions, beliefs, or behaviour are altered or controlled by some form of social communication" [31]. Following this definition, it is hard to think of situations where we can confidently claim we are not influenced in what we think or act. Therefore, it comes as no surprise that social influence is a heavily researched topic, with a long history across diverse disciplines like sociology, social psychology, marketing, economics, media or political studies.

Given the ubiquitousness of social influence processes, the first question one can raise is why do people allow themselves to be influenced? Extant literature suggests that individuals accept influence and engage in the induced behavior in order to fulfill specific needs. Deutsch and Gerard [34] consider there are mainly two types of needs that drive influence processes: the need to be right and the need to be liked. To satisfy the need to be right, individuals accept information from others about the true state of a particular aspect. This is called *informational influence* and is defined as "influence to accept information obtained from another as evidence about reality" [34]. In the same time, to satisfy the need to be liked, individuals adapt to the expectations of others. This is called *normative influence* and is defined as "influence to conform with the positive expectations of another" [34]. While this dichotomy has been accepted to a large extent among marketing scholars, several decades later Cialdini and Goldstein [28] present a more nuanced view. The authors consider that individuals allow to be influenced in order to satisfy one or more of three goals: (1) to form accurate perceptions of reality (goal of accuracy); (2) to develop and preserve meaningful social relationships (goal of affiliation) and (3) to maintain a favorable self-concept. The first two goals can be linked to the informational/normative influence concepts proposed by Deutsch and Gerard [34]. The third goal is related to the individual need for a positive self-evaluation (i.e., the need to feel good about who one is) and implies that an individual decides to follow or reject the attitudes, opinions or behaviors of another based on whether they are consistent with her values.

The goal-directed view described earlier implies that in order to satisfy specific goals, individuals accept to be influenced and behave accordingly. However, the above frameworks do not explain how profound the induced changes are within an individual and how long they will persist. Kelman [74] considers that changes in attitudes and behavior occur at different levels and the difference in levels corresponds to difference in the social influence processes that caused them. The author proposes there are three distinct influence processes: compliance, identification, and internalization. Compliance occurs "when an individual accepts influence because he hopes to achieve a favorable reaction from another person or group" [74]. The individual does not accept the behavior because it is intrinsically rewarding, but because he expects to obtain rewards or avoid punishments by engaging in it. Thus the primary motivation for compliance is the desired social effect. Identification occurs "when an individual accepts influence because he wants to establish or maintain a satisfying self-defining relationship to another person or a group" [74]. In this case, the individual does believe in the behavior induced, but what is particularly important to him is not the actual behavior but the association between him and the others. Thus the primary motivation for accepting influence is given by obtaining or preserving the desired association. Internalization occurs "when an individual accepts influence because the content of the induced behavior (...) is intrinsically rewarding". In this case, the induced behavior is intrinsically rewarding to the individual, and this serves as the primary motivation for accepting influence. Burnkrant [24] shows that each of the three influence processes relates to one of the two influence types proposed by Deutsch and Gerard [34] and that informational influence can lead to internalization while normative influence to compliance and identification. All three processes might lead to the same overt behavior, but the cause and mechanism are different. This has strong implications for understanding the conditions under which the induced behavior will be manifested and how persistent it will be.

At a very high level, most existing literature aims to answer one or several of the following four questions. The first is **who** is influencing or being influenced. This stream is mainly concerned with the characteristics of the source (e.g., influencer) and the target (e.g., influencee) of social influence. The second is **what** is being transmitted. This stream deals with characteristics of the message being transmitted and of the medium through which it is transmitted and typically investigates the characteristics that favor faster / slower transmission. The third is **how** are people influenced. This stream explains the psychological processes that lead individuals to influence or to be influenced. The fourth is **how much** are people being influenced. This stream aims to quantify the effects of social influence on overt behavior.

In the following, we will briefly discuss important results in the marketing literature along the four dimensions. We will focus on recent publications but acknowledge classical work, as well as related literature from other disciplines. The list of articles addressed is by far not exhaustive, as our purpose is not to provide a literature review, but rather to present the main research topics falling under this umbrella. To this end, we will discuss the articles in chronological order, to illustrate the evolution of topics over time. For many articles, it was difficult to categorize them into just one category. In such cases we assigned the articles to the category where we believe their contribution was the strongest.

Who?

One of the earliest works in the field belongs to Katz and Lazarsfeld [73]. When studying the flow of information between the media and the public, Katz and Lazarsfeld noticed that people are influenced in making decision more by other people than by the media. According to the two-flow model of communication the authors developed, a minority of individuals (termed "opinion leaders"), act as intermediaries between the mass media and the rest of the population and thus have a significant effect on their attitudes and behavior. A large body of work that followed this study was dedicated to identifying those individuals, often called opinion leaders or influencers, who have a disproportionate effect on the attitudes, opinions and behavior of many others. A comprehensive review of the work until the 90's across several disciplines can be found in Weimann [139]. In this review, the author proposes that an opinion leader can be described by three characteristics: (1) who one is (personality); (2) what one is (e.g., expert) and (3) strategic network location. As we will see below, most subsequent works characterized opinion leaders according to one or several of these features.

Before diving into the marketing literature, we note that identifying influencers based on strategic network location has been extensively studied by researchers in social networks and network science. A variety of methods have been proposed, including: betweenness centrality [43], eigenvector centrality [21], node accessibility [130], k-shell [79], dynamical influence [80], expected force [83] or collective influence [98].

Turning to recent marketing literature, Goldenberg et al. [52] investigated the role of opinion leaders in new product adoption. Using data from three experimental studies, the authors show that opinion leaders identified as either experts in the field or social leaders (people with a large number of ties) have a different impact on new product adoption. Expert opinion leaders are referred more often for incremental innovations, while social opinion leaders are referred for radical innovations.

In a simulation study, Watts and Dodds [138] question the idea that a few nodes are responsible for the success of large cascades The authors show that except few, rather exceptional cases, influencers identified as central nodes in the social network are not significantly more influential than peripheral ones. The authors test a variety of models and assumptions and show that under most conditions, the success of a cascade is not due to few highly influential individuals starting it, but to a critical mass of individuals who are easily influenced and who influence other, easy-to-influence, individuals.

Goldenberg et al. [51] reach, however, the opposite conclusion. The authors investigate how the hubs in a social network (people with a disproportionately large number of links) influence the aggregate diffusion of innovation process. The authors consider two types of hubs: innovator (hubs who are genuine innovators) and follower (hubs who are not genuine innovators but adopt earlier because of exposure to other adopters). Using an empirical dataset on the adoption of digital items used for customizing home pages on Cyworld (a social network website in South Korea), the authors show there is a positive effect of hubs on diffusion. Innovator hubs drive mainly the speed of adoption while follower hubs drive the total number of people who adopt. Furthermore, the adoption of both types of hubs can be used to predict the entire diffusion.

Nair et al. [101] document as well a significant effect of opinion leaders on behavioral change. Using data on prescription behavior of physicians in the United States, and leveraging a natural experiment: the introduction of new regulations, the authors show that the physicians' prescription behavior is significantly affected by opinion leaders while the behavior of opinion leaders is not affected by non-leaders.

Trusov et al. [131] go beyond the mere network structure and propose a model to identify influential users in online social networks based on the effect they have on the activity level of their friends. Using field data from a social networking website, the authors show that most of the links in the network do not significantly influence user behavior: on average, a user's activity is influenced by only 20 % of her friends. Under these conditions, identifying influencers based on purely structural measures (e.g., degree, betweenness centrality) might not be effective as they cannot distinguish between the different types of links. Therefore, the main contribution of the method over purely structural measures is that it distinguishes the links which affect user behavior from the many other links. In addition, the authors show that personal profile information (e.g., gender), friend counts and profiles views are not predictive of who is influential.

The articles described so far present mixed evidence about the best injection points, if any, in viral marketing campaign. To address this, Hinz et al. [58] conduct two field experiments and a large scale, real-life viral marketing campaign to empirically test four seeding strategies: hubs (high degree), bridges (high betweenness), fringes (low degree) and random selection. The authors find that in all cases, seeding to hubs and bridges has a comparable effect, which significantly outperforms seeding to fringes or at random. Because identifying bridges requires evaluating the betweenness centrality, which is computationally expensive, the authors propose seeding to hubs as the best seeding strategy. However, the authors make a cautionary note and state that seeding to hubs is expected to work only if the objective is to generate awareness as opposed to persuasion as they find that hubs are not more persuasive than the rest of the individuals. If the objective is the latter, the authors recommend the use of demographics and product-related characteristics.

Katona et al. [72] investigate the effect of local network structure among friends who adopted and their characteristics on the adoption likelihood, while controlling for the characteristics of potential adopters. The authors study the growth of a social network, where individuals can only register (i.e., adopt) if they receive an invitation from an existing member. They assume the social network mapped from the website 36 weeks after the registration contains all the real-life relationships between potential adopters and that no relationships were formed as a result of the adoption behavior. Under these two strong assumptions, the authors find there is a positive effect of the proportion of friends who adopted and their local clustering coefficient on the likelihood to adopt, while the average total degree of neighbors who adopted has a negative effect. This implies the more interconnected the friends who adopted, the higher the likelihood to adopt but the higher the degree of an adopter friend, the lower her influence. Furthermore, contrary to Trusov et al. [131], demographic variables were found to have predictive power, with women being more influential than men and younger people more influential than older.

The significant effect of demographic variables was reported as well by Aral and Walker [11] in a large-scale experiment on Facebook regarding the adoption of a new app. The authors found that men are more influential than women, younger users are more susceptible than older users and the least susceptible to influence are married individuals.

Focusing on the network structure, Banerjee et al. [13] propose two measures (communication and diffusion centrality) to identify influencers based on how central they are for spreading information. Using data on the diffusion of a microfinance loan program in 43 Indian villages, the authors show the two measures are strong predictors of the village participation rate in the program.

Hu et al. [63] show that is not only the network position of the influencers or the influencees that has an effect on new product adoption, but also their social status. Using data on the adoption of commercial kits used in genetic treatments by life scientists, the authors show that for products having the potential to boost one's status, there is an inverted U-shape relationship between status (operationalized as degree centrality in the co-authorship network and the total number of publications) and: (1) the likelihood to adopt regardless of social influence (effect explained by middle-status anxiety) and (2) susceptibility to social influence (effect explained by status conformity). Furthermore, the status of an adopter was found to have a positive effect on the adoption likelihood of the peers.

To identify influencers, Chen et al. [26] propose a two-step framework. In the first step, a weighted social network is inferred that encodes the importance of different relationships between actors (e.g., friendship, advice). In the second step, the influencers are identified as the most central individuals in this network (by degree or eigenvector centrality), where the centrality measure takes into account the weighted edges. Similarly to [13], the authors investigate the diffusion of a microfinance product among inhabitants of 43 villages in India and, in addition, the spreading of Super Bowl ads among undergraduate students. They show that influencers identified by taking into account the different relationships between actors are better spreaders than those identified by considering only the binary adjacency matrix. The results show that relationship characteristics have an impact on the diffusion process, which if ignored leads to sub-optimal influencer identification.

Zhang et al. [143] investigate sharing behavior (referred to as rebroadcasting) and consider it is determined by two types of factors: influence-related (who) and content-related (what). By analyzing the tweets of ten well-known business schools, the authors estimate the joint effect of content and influence on rebroadcasting behavior. The results show that in addition to content, the interaction between content and followers (fit between the content and the audience) is an important driver of rebroadcasting behavior. Furthermore, the model outputs a measure of influence and susceptibility that can be used to identify seeding strategies.

Peng et al. [110] consider the characteristics of senders and receivers jointly and investigate how the network overlap among users influences content sharing. By analyzing the tweets of the Fortune 500 companies and the sharing of sponsored ads on Digg, the authors show the number of common friends, common followees, common followers and common mutual followers (users following and followed by both the sender and the receiver) has an effect on the propensity to share. Using a model that allows determining the contribution of individual seeds on influencing a receiver, the authors show that identifying the best senders while taking into account the network overlap is 35-70% more effective than targeting senders with no network overlap with their followers.

Phan et al. [112] study if influence in a network is a consequence of an individual's superior knowledge or expertise. The authors consider two types of individuals: independents - who receive information from outside the network and adopt; and imitators - who receive information from the independents and decide whether to adopt. In a simulation study, the authors found that in several cases, there is an inverse relationship between the influence and knowledge/expertise of an individual. On the one hand, imitators can be more influential than independents because they are more likely to benefit from multiple sources of information. On the other hand, independents can increase their influence by connecting with others to provide better information. However, when the linking process is characterized by high homophily (independents linking only to independents), the influence of the independents decreases as it reduces the amount of information they have access to. Lastly, in noisy communication channels, independents can become more influential than imitators. Based on these results, the authors propose that the best seeders are not the experts themselves but the neighbors of the experts.

Lambrecht et al. [82] investigate the responsiveness of early trend propagators on Twitter to targeted firm ads. Using two field experiments: a campaign ran by a charity to create awareness of homelessness around Christmas and a campaign run by a new fashion label, the authors show that early trend propagators (identified as individuals who post using a keyword or hashtag that is trending that day) of organic trends are less responsive than individuals who post later, while early trend propagators of firm-sponsored trends are as responsive as others. Furthermore, this effect is moderated by how unique and commercial the message is. For messages that are perceived as both more unique and less commercial, early trend propagators are as responsive as others. Drawing on self-determination theory [33], the authors explain the lower responsiveness of early trend propagators by their concern for self-presentation. As they present themselves to the Twitter audience as individuals who are knowledgeable about the latest trends, early trend propagators have little motivation to engage with anything different, like messages from advertisers, as long as this does not help them achieve their goal.

What?

So far we considered which features of the source and target affect social influence processes. In this section, we will discuss the effect of the characteristics of the message and the medium.

Stephen [125] investigate how the structure of the social network influences the generation of new ideas. The authors conduct randomized controlled experiments in which participants need to make decisions about crowdsourcing. By manipulating the network position of the participants, the authors show that a high network clustering has a negative effect on the innovativeness of the ideas generated.

Aral et al. [10] investigate the effect of two viral product features: active personalized referrals and passive broadcast notifications on the adoption of a Facebook app. Using data from a controlled experiment on 1.4 million Facebook users, they find the two viral design features can increase social contagion by up to 400 %, resulting in higher conversion rates compared to traditional ads or paid search campaigns.

Berger et al. [20] investigate the psychological drivers of immediate (soon after consumers learn about the product) and ongoing (in the weeks and months after they first learned about the product) word of mouth (WOM). Contrary to widespread belief, the authors consider that people do not talk more about products that are interesting but about products that come easier to their mind. Using an empirical study of face-to-face WOM campaigns for more than 300 products together with a field and a lab experiment, the authors show that products that are more publicly visible or cued more frequently by the environment are more accessible and in consequence stimulate more immediate, ongoing and overall WOM compared to just interesting products. The authors furthermore investigate which promotional actions from the company result in higher WOM. They find that offering the product itself or nonproduct extras (e.g., logo hats, recipes) has a positive effect on WOM while offering samples, discounts or coupons does not have a significant effect.

Schulze et al. [119] investigate the success of diffusion for low versus high utilitarian products. Using data on the installation of 751 Facebook apps, the authors show that consumers react differently to viral marketing campaigns for low versus high utilitarian products in a social network that is fun-oriented. The authors consider four types of sharing mechanisms: unsolicited messages, messages with incentives, direct messages from friends (one to one), broadcast messages from strangers (one to many) and show that their effect on the reach of the app is moderated by how utilitarian the product is. In consequence, when trying to diffuse primarily utilitarian products, companies should not rely on the same sharing mechanisms as for less utilitarian product as using the same strategy may result in fewer installations that not using any viral marketing campaign at all.

Packard et al. [108] study the effect of implicit ("speaker's declaration of his or her own tastes") versus explicit ("a declaration that the speaker finds the object appropriate for an audience") endorsement styles on WOM impact. Using observational and experimental data on product reviews across different product categories, the authors show that explicit endorsements (e.g., "I recommend it") have a larger effect on both how much consumers think they will like the product and their willingness to purchase it than implicit ones (e.g., "I like it"). Explicit endorsements are more persuasive not only because the senders are perceived to have liked the product more, but also because they are perceived as having higher expertise. In contrast, less knowledgeable consumers were found to use more often explicit than implicit recommendations as they are less aware of the preference heterogeneity. Thus situations can arise when receivers are persuaded more by less knowledgeable consumers.

How?

So far we have seen which characteristics of the sender/receiver and the message/medium have an effect of the influence process. In this section, we will discuss the psychological motivation of these and other social influence effects.

Dholakia et al. [35] investigate what drives individuals to participate in virtual communities. The authors consider two types of reasons: individual based (purposive value, self discovery, maintaining interpersonal connectivity, social enhancement, entertainment value) and group based (group norms and social identity). Using survey data, the authors show the individual reasons are antecedents to group reasons, which in turn influence participation. Moreover, the authors find the effect of individual reasons is moderated by the type of community: network based vs. small group based.

Algesheimer et al. [8] study how customers' intentions and behavior are influenced by their relationship with the brand community. By conducting a survey among members of car clubs in Switzerland, Austria and Germany, the authors show that: (1) brand relationship quality leads to stronger identification with the brand community; (2) identification with the brand community leads to greater community engagement and lower normative community pressure; (3) greater community engagement leads to higher normative pressure; (4) the intentions to continue membership, actively participate in the community and recommend the brand are positively influenced by increasing community engagement and negatively influenced by increasing normative pressure; (5) greater normative pressure leads in addition to reactance, which has a negative effect on brand loyalty intentions and brand membership continuation; (6) all described intentions lead to subsequent behaviors.

Sridhar et al. [124] investigate how online ratings of other consumers moderate the effect of personal experience with the product on consumer's rating. The authors consider four types of experiences: positive features, regular negative features (i.e., acceptable negative features), product failure (i.e., unacceptable negative features) and product recovery. By analyzing online reviews of hotels in Boston and Honolulu, the authors show that when consumers have a positive experience, the higher the online ratings, the weaker is the effect of positive features as individuals feel a high need for uniqueness. Furthermore, the higher the ratings of other consumers, the weaker the negative effect of regular negative features and the stronger the negative effect of product failure on consumer's rating. The results highlight the bi-directional nature of social influence. The reviewer influences others, but at the same time is influenced by others in writing the review and deciding on the rating value.

Luo et al. [93] investigate the effect of social influence on group buying (i.e., the purchase of discounted products/services on websites like Groupon). The authors suggest group buying can be seen as a two-step process. In the first step, a decision is made if to buy or not. Conditional on a positive decision to buy, in the second step a decision is made when to redeem the deal. The authors investigate how deal popularity affects both decisions. Using observational data from a Groupon like company, the authors show that deal popularity affects both the decision to buy (first step) and the decision to redeem (second step) and this effect is moderated by referral intensity and group consumption.

Hamilton et al. [55] investigate how the content of early replies in online forums affect the content of subsequent replies. By analyzing discussion threads on Tripadvisor and DISboards.com, the authors show that if respondents view previous answers to a question, they tend to focus more on the attributes mentioned in the answers and not on the ones in the question. This effect is caused by the individual's desire for affiliation and can lead to suboptimal or incomplete information provided to the person asking the question.

Zhang et al. [142] investigate the effect of social connections on individual goal attainment and spending. Using data from two observational studies on online gaming markets and a controlled experiment, the authors show there is a nonlinear effect of social connections and interactions on goal attainment and spending behavior. With increasing number of connections/interactions, after a certain point information overload occurs and the utility of the information received decreases. Individuals benefit from new information up to a point after which the utility of the information received decreases due to high information processing costs. Furthermore, the authors show this effect is moderated by individual experience. Individuals with low experience benefit more from social connections/interactions but also have stronger negative marginal effects from too many interactions compared to individuals with high experience.

Morvinsky et al. [99] show that customer uncertainty about the product and homophily with current adopters moderate the effect of adoption stock on new product trial. Using data from controlled experiments and a field experiment on energy drink choice, the authors show that a large adoption stock has a positive effect on the trial probability only under moderate uncertainty and high homophily. No effect was found for low uncertainty (at any level of homophily), while a negative effect for high uncertainty and high homophily. The results bring evidence against common practice that signaling a large stock of adoption to potential customers always has a positive effect.

Meyners et al. [97] study the effect of geographic proximity on social influence. Using an empirical dataset on the adoption of a mobile phone provider and controlled experiments, the authors show that geographic proximity leads to higher perceived homophily, which in turn leads to stronger social influence and this effect increases with decreasing tie strength.

How much?

So far we have discussed which characteristics of the source/target/message/medium and which psychological processes drive social influence. In this section, we will present work that quantifies the effect of such characteristics and processes on a behavioral outcome of interest.

Chevalier and Mayzlin [27] study the effect of book reviews on sales. By comparing sales ranks of books across two online retailers: Amazon and Barnes and Noble, the authors show that the improvement in book reviews on one site leads to more sales on that site relative to the other. Furthermore, it was found that most reviews are positive on both sites and that a negative review has a stronger effect in decreasing sales than a positive review in increasing sales.

Godes and Mayzlin [50] investigate the effect of firm created (i.e., exogenous) word of mouth on sales. Using data from a field test on the WOM marketing campaign of a restaurant chain and a controlled experiment on website usage, the authors show that firm created word of mouth has a positive effect on sales. Contrary to common intuition, the authors show that for products with low initial awareness, the best disseminators of information are not the loyal but the less loyal customers. This effect is explained by the "strength of weak ties" theory [53]: the friends of loyal customers are likely to know about the product existence already before the start of the WOM campaign. Furthermore, the authors show the customer loyalty of opinion leaders moderates the effect of opinion leadership on WOM creation. Opinion leaders are effective spreaders only if they are loyal customers. They need to have extensive experience with the product, or they risk losing their status if they become too innovative.

Luo et al. [92] investigate the long-term financial impact of negative word of mouth on firm security prices. Using an observational dataset of complaints voiced by consumers against airline companies in the United States, the authors show that negative word of mouth has a significant negative effect on the firm's cash flows and stock returns, which in turn creates more negative word of mouth and this negative effect increases in the presence of competition.

Algesheimer et al. [7] study the effect of community participation on customer behavior, while controlling for self-selection. Using a field experiment conducted among eBay users, the authors show that: (1) email invitations significantly increase participation and (2) community participation has no effect on the number of bids placed and revenue earned and a negative effect on the number of items listed and the amount spent. The results are explained by community participation having an educational value for community members: they become more selective and efficient sellers and more conservative in their spendings. Importantly, to illustrate the self-selection problem specific to many social influence studies, the authors compare their results against a model where they do not control for customer self-selection to participate in the community. The results are in stark contrast with the original model, showing a positive effect of community participation on all four variables of interest: number of bids placed; revenue earned, number of items listed and amount spent. The authors consider that the positive effect of community participation on engagement with the firm commonly reported in previous literature might be caused by the impossibility to account for self-selection.

Du and Kamakura [36] investigate if social contagion exists as well for consumer packaged goods (i.e., products that customers use and replace frequently). Using data on purchase history for 67 newly introduced products, the authors show there is a positive effect of the previous adopters on the trial probability.

Iyengar et al. [65] investigate how social contagion and opinion leadership affect the diffusion of new products. Using data on prescription behavior among physicians in the United States, the authors show that contagion exists after controlling for potential confounders like marketing efforts. The contagion effect is moderated by the source's volume of product usage (higher usage leading to higher contagion) and the target's perception of her opinion leadership (higher perception leading to lower susceptibility). Furthermore, the authors find there is a low correlation between the most two common operationalizations of opinion leadership: sociometric and self-reported, which implies they might capture different constructs. The former tend to adopt earlier while the latter tend to be less susceptible to influence from their contacts.

Nitzan and Libai [104] study the effect of social influence on customer retention. Using data from a telecommunications operator, the authors show that exposure to defecting neighbors leads to 80% increase in the defection hazard, after controlling for potential confounders.

Risselada et al. [115] investigate the dynamic effect of recent and cumulative adoptions on the adoption of high technology products while controlling for direct marketing. Using data on smartphone adoption in a social network (derived based on called detailed records), the authors show that cumulative adoptions and direct marketing have a positive, time-decreasing effect on the time of adoption (nr of months between the introduction of the mobile phone by the telecommunications operator and its adoption), while the effect of recent adoptions is also positive but constant over time.

While most work described so far documented the effect of social influence on the trial probability of a product (e.g., new product adoption), Iyengar et al. [64] take a step further and investigate the effect of social influence on both trial and repeated usage. Using data on prescription behavior among physicians in the United States, the authors show that contagion is at work in both stages and that who is most influential and most susceptible is different for trial and repeat. The immediate colleagues are influential in both trial and repeat, while physicians with high centrality in the discussion and referral network are influential only in trial, and only among physicians with low self-reported opinion leadership. The most susceptible in trial are the physicians who do not consider themselves opinion leaders (low self-reported opinion leadership) while in repeat are those in the middle of the status distribution, as measured by degree centrality. The authors explain the results by informational social influence reducing risk in trial and normative social influence increasing conformity in repeat.

Chae et al. [25] study the spillover effects of seeding campaigns beyond the focal product. Using a dataset of 390 campaigns for cosmetic products on Naver (one of the largest Internet portals in South Korea), the authors find that seeding campaigns have a positive effect on the number of mentions of the product among unseeded consumers and a negative effect on the number of mentions of other products belonging to the same brand and of products from competing brands that are in the same category.

Fossen et al. [42] investigate the effect of television advertising on online WOM. Using data on television advertising and minute-by-minute Twitter conversations about brands and programs, the authors show advertisements have a positive effect on online mentions for both brands and programs, but advertising in programs that have higher than expected online activity does not lead to an increase in WOM for the advertised brand.

Seiler et al. [120] investigate the effect of microblogging activity on TV consumption. By leveraging a natural experiment: a 3-day block of the comments functionality on Sina Weibo due to political events, the authors show the effect of WOM (measured as the elasticity of TV ratings with respect to comments posted) is considerably lower than previously reported in the literature. The authors argue the upward bias in previous work is caused by the limited ability to control for endogenous effects in non-experimental studies. Furthermore, WOM before the show was not found to have an effect, while WOM after the show had a significant effect on TV ratings, for both positive and negative comments. This implies that WOM affects TV viewership not by informing or persuading people as it would be expected, but by providing a complementary activity. Taking part in controversial debates raises the consumers' utility from watching the show.

In this thesis we aim to bring a contribution to several streams as follows. In Chapter 3 we contribute to literature on influencer identification ("Who") by proposing a method to identify influencers based on the reaction of the social group to their actions. In Chapter 4 we contribute to literature quantifying the effects of social influence ("How much") by providing a new conceptualization of social influence and illustrating its effect on repurchase behavior in the presence of variety seeking. In Chapter 5 we bring an interdisciplinary contribution by extending methods for influencer identification to solve complex policy problems.

Chapter 3

Identification of influencers through the wisdom of crowds

Identifying individuals who are influential in diffusing information, ideas or products in a population remains a challenging problem. Most extant work can be abstracted by a process in which researchers first decide which features describe an influencer and then identify them as the individuals with the highest values of these features. This makes the identification dependent on the relevance of the selected features. Furthermore, most work was developed for cross-sectional or time-aggregated datasets, where the time-evolution of influence processes cannot be observed. We show that mapping the influencer identification to a wisdom of crowds problem overcomes these limitations. We present a framework in which the individuals in a social group repeatedly evaluate the contribution of other members according to what they perceive as valuable and not according to predefined features. We propose a method to aggregate the individual evaluations into a collective judgement that considers the temporal variation of influence processes. Using data from three large news providers, we show that the members of the group surprisingly agree on who are the influential individuals. The aggregation method addresses different sources of heterogeneity encountered in social systems and leads to results that are easily interpretable and comparable within and across systems. [1]

Keywords: influencer identification | wisdom of crowds | computational social science

[1]**Author Statement:** This work is in collaboration with Claudio J. Tessone and Rene Algesheimer from the University of Zurich, Switzerland. The article is published as Tanase R, Tessone CJ, Algesheimer R (2018). Identification of influencers through the wisdom of crowds. *PLoS ONE* 13(7): e0200109. https://doi.org/10.1371/journal.pone.0200109. The work was supported by the University Research Priority Program "Social Networks" of the University of Zurich. We would like to thank Patrik Schilter for assistance with data collection and the S3IT Team at the University of Zurich for support with processing data.

3.1 Introduction

Firms, political parties and organisations increasingly rely on engineering social contagion to spread products, ideas or behaviours. Already for more than half of a century researchers have realised that a relatively small number of people can have a great impact on the opinions and behaviour of many others. The concept of opinion leaders (influencers or influentials[96, 139]) was first introduced by Katz [73] in the study of the two step model of communication flow between the mass media and the public and since then it has been revisited in a plethora of studies across many academic disciplines [29, 138, 51, 131, 65, 58, 11, 13, 77]. Extensive research has shown that influencers drive new product adoption [29, 51, 65, 58], public health policies [77] or voting behaviour [139]. In consequence, a large body of literature has been devoted to the identification of influencers [21, 43, 75, 79, 130, 80, 83, 98, 131], which is still considered today as one of the most important and challenging problems [131, 98, 77].

In general, influencers can be described by a combination of three factors: personification of values (who one is), competence (what one knows) and strategic network location [139, 51]. Most existing identification methods are constructed by selecting one or several features belonging to these factors and identifying the individuals with the highest values of these features. Such features range from psychological traits [139] to expertise [84] or position in the social network (e.g. betweenness centrality [43], eigenvector centrality [21], node accessibility [130], k-shell [79], dynamical influence [80], expected force [83] or collective influence [98]). An important limitation of this kind of approach is that the selection of relevant features is done a-priori by the researcher or practitioner, according to his own subjective preferences and thus the identification of influencers strongly relies on the assumed relevance of the selected features. Hinz et al. [58] have shown that influencers identified as individuals with either high degree or betweenness centrality are better spreaders of information than individuals with low degree. On the other hand, Watts et al. [138] have shown that except few, rather uncommon cases, influencers identified as central nodes in the social network are not significantly more influential than peripheral ones. This evidence against an universal set of features describing influencers can be explained by the complexity of the influence process. Personal influence has been shown to operate through several latent mechanisms (e.g. contact, socialisation, status competition, social norms) which have a different impact across the five stages of the decision process (knowledge, persuasion, decision, implementation, confirmation) [116, 102]. Under these circumstances, selecting a set of features that describe influencers is difficult without detailed knowledge of the context in which the influence process takes place. Furthermore, most methods can only be applied to a time-aggregated dataset [43, 21, 130, 79, 80, 98, 83], which neglects the inherent temporal nature of the influence relationships. There exist several attempts to extend methods to the temporal case [59], but the problem is far from being solved. The widespread belief is that adding a temporal layer leads to much more complex objects, whose study requires the development of sophisticated tools [121, 81].

In this article, we show that mapping the influencer identification problem into a wisdom of crowds one overcomes these limitations. The wisdom of crowds phenomenon [127] was first described by Galton [44] when he observed that social groups can make more accurate collective judgements than expert individuals [44, 127]. Since then, this phenomenon has raised great interest among both researchers and practitioners. People have been shown to make surprisingly accurate judgements when their opinions are aggregated and this concept has been applied to solve a large variety of problems, from prediction markets to informed policy making [127]. The idea also made its way into mainstream applications, being an important mechanism behind creating content on social information sites such as Wikipedia, Quora or Stackoverflow. We present a framework in which the individuals in a social group repeatedly evaluate the contribution of other members according to what they perceive as valuable and propose a method to aggregate the individual evaluations into a collective judgement. In doing so, we do not make any assumption on what are the relevant features of the influencers, but we let each individual decide on his own, based on the preferences and beliefs held at that point in time. Furthermore, we show that under this mapping, the temporality of the data provides in fact a simplification of the influencer identification. This supports a recent study [121] which shows that temporal complexity may in fact simplify certain problems if seen through the right perspective.

3.2 Results

Identification approach

We consider that individuals become influential due to a latent construct they possess which reflects their knowledge and skills, as well as preferences and beliefs. We call this unobserved construct the latent potential to influence. This potential is revealed during social interactions and can be evaluated by other participants through a voting system (e.g. up-votes on discussion platforms). While traditional methods use features set *a priori* by the practitioner or the scientist, our method uses the crowd's judgement, expressed through votes. Operationalising influence in terms of votes reflects both the heterogeneity in skills and knowledge between contributors and the heterogeneity in preferences and beliefs between the evaluators. The latent potential to influence is uncovered by aggregating the individual evaluations. Commonly used methods include the total number of positive evaluations (variations of this are commonly used in social information sites), the mean or the median [44]. However, when applied to systems characterised by a heavy-tailed distribution of variables describing the system (like many social media platforms), such methods might be biased as the quantities they aggregate are not directly comparable. To address all previous shortcomings, we develop the *influence potential (IP)*, a new aggregation method.

In the remainder of the article we will use the term *event* to describe a time-window capturing social interactions between individuals. Without loss of generality, the events take place at different points in time, which implies there is a sense of temporality in the data. However, this assumption is not restrictive and the events can also be concurrent. For every event we rank all participants in increasing order of votes received and compute the event rank of an individual in an event as the rank normalised by the total number of participants in the event plus a constant. Formally, the *event rank* of individual i in event t is defined as $R_t(i) = \mathrm{rank}(i)/(n_t + c)$, where n_t represents the number of participants in event t (*event size* of t) and c is an additive constant which controls for the event size. Further, let $\mathscr{E}^i$ be the set of events where i participated. The influence potential (IP) of an individual is the mean normalised rank over the events where he participated minus the respective variance. That is, the influence potential of individual i is

$$\mathscr{I}(i) = \langle R_t(i) \rangle_{t \in \mathscr{E}^i} - \left(\left\langle R_t(i)^2 \right\rangle_{t \in \mathscr{E}^i} - \langle R_t(i) \rangle^2_{t \in \mathscr{E}^i} \right). \tag{3.1}$$

The variance term is introduced to penalise the lack of consistency in the ranks obtained. The IP reflects the extent to which most participants in an event consistently appreciated the contribution of the individual each time he was active. Notice that we do not impose a criteria on how the contribution should be evaluated. The IP is bounded in the interval $[0,1]$ (see Proposition 1). A value close to zero is obtained for individuals who either consistently rank low in the votes distribution or have a high variation in the votes score across all the events they participated in. Such individuals have a low potential to influence as, either their contribution is rarely appreciated or this happens with a high level of uncertainty, questioning their inherent abilities. On the other hand, a value close to one can only be obtained for individuals who always collect the most votes in the events they participate. Such individuals have a high potential to influence as, due to some construct we do not directly observe, they always attract the highest evaluation. An implicit assumption made in Equation 3.1 is that the activity of an individual (defined as the the number of events attended) is not alone informative for the latent potential to influence but is rather an opportunity for the latent potential to influence to be manifested. In the following, we show using a large dataset of different large-scale news providers that: (1) the members of the social group surprisingly agree on who are the influential individuals, and (2) the extent to which the members agree varies across discussion topics.

3.2.1 Data collection

We collected the complete history of online discussions over a long period of time from three large news providers: CNN, The Atlantic and The Telegraph. Such platforms offer an interactive environment in which users have the possibility to express their views, engage in discussions and possibly shape other's view on the topic. Registered users can post comments in discussion threads and, at the same time, react and evaluate the quality of the posts through a voting tool provided by the platform. The default ordering of the posts on the platform is determined by the number of votes received. The discussions cover a broad range of topics, each thread belonging to one topic category which defines the overall topic of the discussion (e.g. politics, business, etc.). The categories are defined by the news providers and are directly available on the website. Discussion threads for which it was not possible to identify the category have been omitted from the analysis. All platforms are comparable in terms of user experience as they are based on the same technology, provided by Disqus. An overview of the three datasets can be found in Table 3.1 (approximative figures) and a detailed description of the categories in Tables A.1-A.3 in Appendix A.4. In our terminology, the discussion threads represent the events, the contribution of an individual in an event is defined by the total number of posts made in the thread, and the evaluation of the contribution is defined by the number of up-votes received by all posts made in the thread.

Table 3.1 Overview of data analyzed

Dataset	Period	Active users	Threads	Posts	Cat.
CNN	2012 - 2014	9×10^6	3×10^4	23×10^6	13
Atlantic	2007 - 2016	3×10^6	5×10^4	5×10^6	17
Telegraph	2006 - 2016	5×10^6	33×10^4	22×10^6	15

3.2.2 Identification of influencers

We investigate if for each topic there are individuals who consistently receive most votes each time they are active. In the remainder of the article we use $c = 1$ and consider only individuals who participate in at least 10 events. In Fig. A.1 (Appendix A.4) we show the results are robust to the choice of c and later in the article we show the IP is robust to the number of events observed per individual. In Fig. 3.1, upper panels, we show the relationship between the mean event rank (x axis) and the corresponding variance (y axis). It can be seen there are several individuals with a high mean event rank and a low variance (illustrated with dark blue colour code). In consequence, in the lower panels of Fig. 3.1 we observe a heavy tailed distribution of the IP, with several individuals having high values. This is consistent with existing literature, which states that there are just a few influencers compared to the entire population [138]. This result is rather surprising as we would expect a high disagreement between the participants in an event because what is a valuable contribution is decided by each individual on his own, based on his own preferences and beliefs. We later consider two parsimonious mechanisms and show that none can completely explain the results.

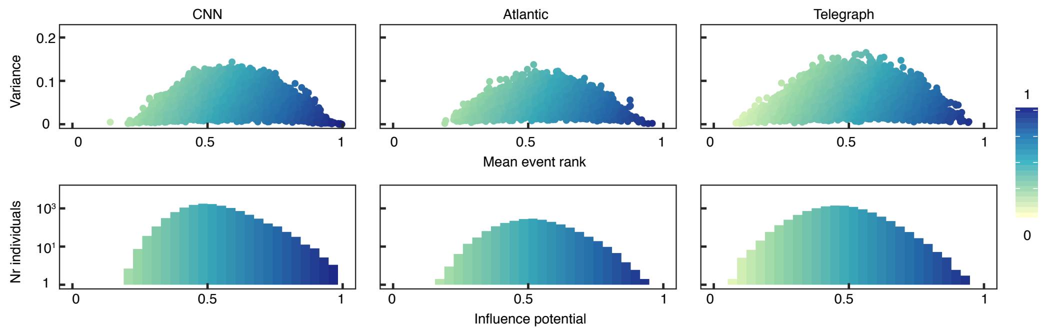


Fig. 3.1 **Identification of influencers**. Data is pooled from all categories. An individual can be described by multiple data points, each being related to his performance in one category. *Upper Panels*: Relationship between the mean event rank (x axis) and the corresponding variance (y axis). There is an inverted U-shape relationship between the mean and the variance of the event ranks. The colour of the points is given by the IP. The individuals with high mean and a low variance have the highest IP. *Lower Panels*: Distribution of the IP.

3.2.3 Zooming in topic categories

We now investigate how the nomination of influencers varies across the topic categories. By doing so, we are able to identify category influencers. Figure 3.2 contains a boxplot of the highest 100 IP scores within each category. To ease the representation, for each dataset we selected the top 10 categories with the highest number of users. We re-labelled each category according to its ranking in terms of number of users among the categories within the same dataset, C1 representing the highest. The list of abbreviations together with the number of users in each category can be found in Tables A.1-A.3 in Appendix A.4. Figure 3.2 shows there is a considerable difference between the highest influencer scores across the different categories ($p < 10^{-16}$, ANOVA test). For example, on the CNN platform, in the first five categories (C1-C5: world, us, opinion,politics, justice) the influencer scores of the top 100 individuals are considerably higher than in the following five (C6-C10: showbiz, tech, health, travel, living). This has practical implications for designing intervention campaigns based on targeting influencers as it shows that in the same system, the extent to which individuals agree on who and what is influential might change depending on the context. Very often in literature it is considered that the extent to which an individual is influential is determined by one or several feature he possesses [139, 51] and it is neglected how, for the same individual, the impact of these features on his perceived influence can vary across different settings. In Figure 3.3 we selected the top 10 individuals with the highest IP in each category and plotted their IP scores for all categories in the dataset. If an individual did not participate in a category, it was represented by a blank cell in the figure. It can be seen that: (1) most individuals participate in very few categories and (2) individuals who participate in more categories have high IP scores only in few. This suggests that individuals who are influential across topics are hard to find, possibly because an important component of the latent potential to influence is the *topic expertise* [139]. This finding is in line with early studies which showed that opinion leadership is topic dependent, with different degrees of overlap between the topics [78]. However, in recent studies this is very often neglected as influencers are mostly identified based only on one (often structural) feature [43, 21, 130, 79, 80, 98, 83]. Targeting for example a well connected individual who is expert in politics to promote a healthy behaviour has a high risk to fail.

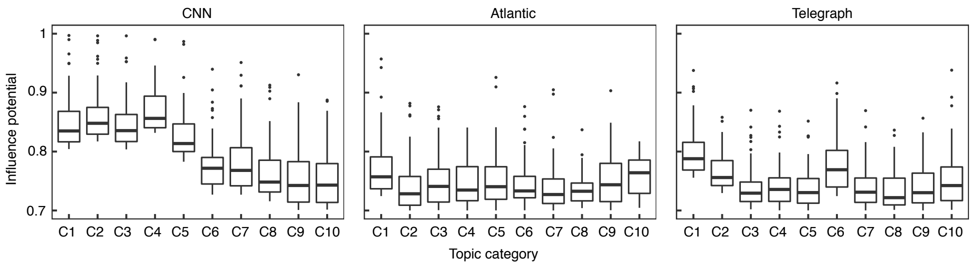


Fig. 3.2 **Influencers within topic categories**. The x axis represents the topic category. The y axis represents the IP scores of the top 100 individuals with the highest IP within the category. The categories are ordered by the number of users.

3.2.4 Different aggregation methods

We compare our aggregation method against three alternatives often encountered in research or practice: the total number of votes (used regularly on social information sites to rank users), the mean and median [44] number of votes. For every topic, we rank all users according to

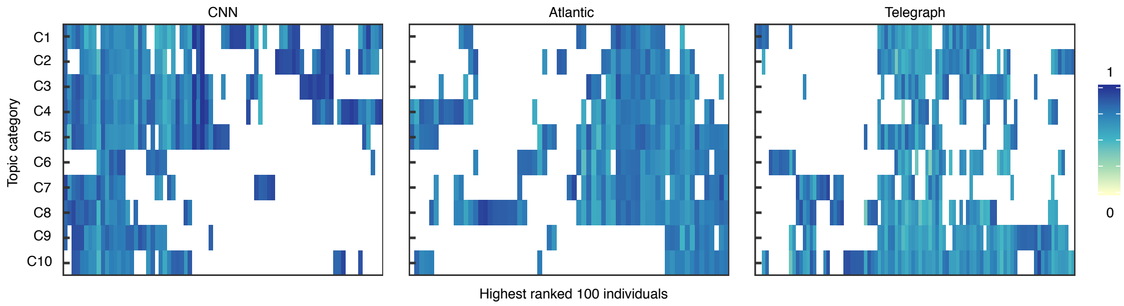


Fig. 3.3 **Influencers across categories**. The x axis represents the IP scores of the top 10 individuals with the highest IP from each category. The y axis represents the topic category. If an individual did not participate in a category, it is represented by a blank cell. Most individuals participate in few categories. Individuals who participate in more categories have high IP scores in only few.

the four methods and calculate the degree of overlap between the highest ranked users. Fig. 3.4 shows the results. The total number of votes leads to considerably different results, with the lowest overlap with the other methods. One reason is that this method does not control for the difference in activity between individuals nor for the difference in size between events. On the other hand, the highest similarity can be observed between the mean the median. Both methods control for the difference in activity between individuals, but not for the event size. In addition, the median is not sensitive to extreme evaluations which can explain the higher difference observed in the CNN dataset. The IP is closest to the median, with a significant but not high overlap between the two.

Compared to existing aggregation methods, the IP has several appealing features. First, it addresses different sources of heterogeneity often encountered in social systems. A predominant characteristic of most social systems (including news platforms) is that there is a heavy tailed distribution of the variables describing the system. Fig. A.2-A.4 (Appendix) show there is a large difference in the number of participants in the events. As the total number of votes in an event is proportional to the event size (Fig.A.5 in the Appendix), it implies that we cannot directly compare the number of votes received in events of different sizes. An aggregation method which does so, like the mean or the total number of votes, could be biased towards participation in large events. By aggregating instead the normalised ranks, we ensure the evaluations are comparable across events. The number of events attended by an individual is as well described by a heavy-tailed distribution (Fig. A.6-A.8, Appendix). This implies that comparing users in terms of the total number of votes received, as it is done on most social information sites, will favour individuals who are very active. To infer the latent potential to influence our approach does not take into account the number of events attended (once a minimum number has been achieved). In this way we are, to some extent, separating the tendency of individuals to be active from their latent potential to influence, making the influencer scores comparable across individuals with different levels of activity. Second, the aggregation method we propose provides normalised results, that are easy to interpret and compare within and across systems. Individuals who are influential have IP scores close to one, while non-influential individuals have scores close to zero. Because of this, the extent to which somebody is influential can be directly inferred from his IP score, without the need of additional information about the system, like it is the case with the other aggregation methods.

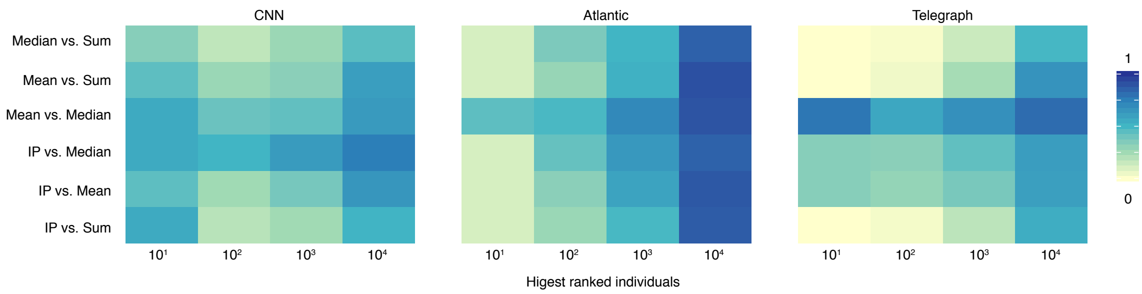


Fig. 3.4 **Comparison of results under different aggregation methods**. We compare the overlap between the highest ranked individuals by different methods. The x axis represents the number of highest ranked individuals. The y axis represents the overlap between the highest ranked individuals by two methods. Data is pooled from all topic categories. The mean and the median are the most similar. The IP is closest to the median.

3.2.5 Robustness checks

Already more that half a century ago, Bass [17] has observed a high correlation between the time a person spends talking and her perceived leadership in the social group. When data is generated by such a mechanism, high IP scores merely reflect *talkativeness* (here defined as the tendency of individuals to post excessively), which is then considered as the main component of the latent potential to influence. To test if data can be explained by the talkativeness effect, we create a null model in which the observed number of votes is uniformly distributed across all posts in a thread. Specifically, for every thread we sample with replacement from all posts a number of times equal to the observed number of votes in the thread. Then we compute the IP as described above, using the sum of randomised votes as input. The procedure is repeated 100 times and the IP under the null model is computed as the mean IP over the repetitions. Under this model, the event rank of an individual in a thread is proportional to his number of posts. Individuals who write more have a higher chance to obtain a high event rank, and thus a high IP. Fig. A.9 in the

Appendix shows that this mechanism can lead to the emergence of individuals with a high IP, even though the allocation of votes is done at random. However, there is little overlap between the highest ranked individuals identified under the two conditions. For the highest ranked 1000 individuals by either IP or the IP under the null model, there is a low Pearson correlation between the scores (cnn: -0.09, atlantic: 0.15, telegraph: 0.23). This implies that in nominating influencers, the crowd considers more (or different) features than just the active participation in the event.

Many social systems are characterised by a rich-get-richer effect [140, 122, 15], where individuals who enter the system early have an advantage over those who enter late. This effect is particularly important on many online discussion forums, including our data source, where the default ordering of the posts on the platform is determined by the number of votes received. This is a case when social influence can undermine the wisdom of crowds [89, 19]. Under such a mechanism, high IP scores reflect the ability of an individual to enter the system early and gain initial votes. To test if data can be explained by the rich-get-richer effect, we create a null model in which a vote is allocated with probability α to a post selected at random and with probability $1-\alpha$ to a post selected according to a preferential attachment model in which posts with more votes are more likely to be selected (see Appendix A.3). For $\alpha \in \{0, 0.1, \ldots, 0.9\}$, for all three datasets the Spearman correlation coefficient between the scores of the highest ranked 1000 individuals by either IP or the IP under the preferential attachment null model is low ($r <= 0.22$). This implies that also preferential attachment is not enough to explain the results and that the individuals identified by the crowd have unobserved features that allow them to obtain the most votes each time they are active.

One concern that might be raised is that high IP values are favoured by participation in a low number of events, even after imposing a threshold on the minimum number of events attended. We show our results are robust to the sample size using the following procedure. We define a sequence of percentiles $q \in [0, q_{max}]$ and for each individual i define a set of events $\mathscr{E}^i_{-q}$ that is constructed by removing at random $q\%$ of all the events where i participated within the topic category. Then compute the IP of i over only the events in $\mathscr{E}^i_{-q}$. To ensure the IP is always computed using at least 10 events we remove from the analysis all individuals who did not participate in more than 20 events within a topic category. Figure 3.5 shows that on average, reducing the sample size at random by even 50% ($q_{max} = 0.5$) does not produce higher IP values. This shows that, once a minimum number of events was observed, smaller samples do not lead to higher IP values.

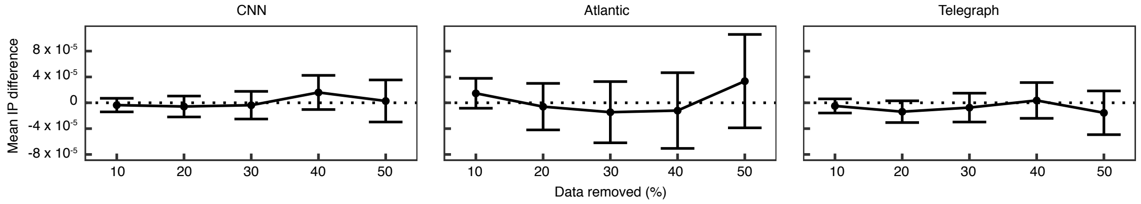


Fig. 3.5 **IP scaling with sample size**. The x axis represents the percentage of events removed at random. The y axis represents the mean difference between the IP scores based on the entire sample and the IP scores based on the random subset. Decreasing the sample size does not have a significant effect on the IP.

3.3 Discussion

Political parties, companies or health organisations are interested to identify influencers and use them as superspreaders of products, ideas or behaviors in intervention campaigns [77, 13, 132]. The dominant mindset is to first identify a set of features that could best describe an influencer and then look for individuals with high values of these features. While this is a perfectly feasible approach, with a high success across a wide range of applications [77, 65, 51], it also suffers from several drawbacks. First, it is limited by itself, as by construction it can only identify individuals with high values of the selected features, irrespective if these are relevant in the given scenario or not. There is no agreement in the literature on which is the best set of features, with many studies challenging previous findings [77, 138, 58]. Furthermore, we believe the importance of these features is both time and context dependent. Who we consider as a reliable source of information might change depending on when we intend to make the decision or its perceived level of risk [77, 38]. Nowadays datasets are much richer than before, with high time-resolution and detailed individual information being frequently the norm. Classical methods were developed to deal with cross-sectional data, as often researchers and practitioners had a single data snapshot available. There are many attempts to extend these methods to account for increasing levels of complexity like temporal variation, but most often this is not straightforward, leading to complicated mathematical descriptions that are computationally expensive or which come at the price of stronger assumptions, making it difficult to apply them in real-time environments. We follow a recently proposed path [121, 81] and show that more complex information can actually simplify the analysis if seen trough the appropriate lens. Influencer identification in temporal systems with a measurable outcome of social interactions (e.g. social media platforms) can be mapped to a wisdom of crowds problem, where individuals decide on their own what is relevant for them at every point in time. By aggregating the individual evaluations, it is straightforward to reveal who is consistently the most influential each time he is active. In our attempt to keep the aggregation method simple and intuitive, we did not consider that the evaluation received by an individual might be influenced by the individuals against whom he is competing. In discussion threads where many influencers participate, it might be more difficult to obtain a high evaluation due to competition dynamics. An extension of the method to account for such cases might provide a valuable contribution. We conclude by mentioning that the applicability of the aggregation method is not restricted to the wisdom of crowds scenario. In particular, it could be applied to quantify performance in any temporal system where a performance metric is measured over time. It could be used in diverse disciplines like network science to quantify centrality in temporal networks, management to quantify performance of employees or sports to identify the most valuable players.

Chapter 4

The effects of social influence and variety seeking on repurchase behavior

Understanding how people influence or are influenced by their peers can help us understand the flow of market trends, product adoption and diffusion processes. Most existing work on social influence considers change in purchase behavior as a dependent variable and thus an individual is influenced if she was determined to change her behavior. However, nowadays people are faced with countless buying options, thus repeatedly purchasing the same product can be considered the exception rather than the norm. In this paper, the authors propose a theoretical framework in which the decision to repurchase or switch to a new product is related to two types of stimuli: intrapersonal (related to variety seeking behavior) and interpersonal (related to exposure to the social group). Using data from two controlled experiments and an empirical study, the authors show that (1) variety seeking has a negative effect on the probability to repurchase; (2) social exposure supporting the repurchase has a positive effect and (3) social exposure supporting switching to a new product has a negative effect. Furthermore, they construct a theoretical model of product choice and show that: (1) when both variety seeking and social exposure have a positive valence towards product switch, the effect of social exposure on product switch is overestimated as the decision to switch is already made by the consumer and (2) when variety seeking has a positive valence towards product switch and social exposure has a positive valence towards repurchase, a repurchase reflects an actual influence effect. [1]

Keywords: social influence | variety seeking | repurchase behavior

[1]**Author Statement:** This work is in collaboration with René Algesheimer (University of Zurich, Switzerland) and Utpal Dholakia (Rice University, USA). It was supported by the University Research Priority Program "Social Networks" of the University of Zurich and by the Jones Graduate School of Business, Rice University.

4.1 Introduction

Governments and organizations increasingly rely on engineering social contagion to spread products, ideas or behaviors. The key idea behind social contagion is that our opinions and behavior are to a large extent influenced by individuals in our environment or social network [73, 30, 52, 138, 51, 131, 65, 58, 11, 77]. In the majority of studies, the concept of social influence is associated with opinion or behavioral change. This (implicit or explicit) assumption is made either in the definition of influence itself or in evaluating its effect on a process of interest. When they introduced the concept of opinion leader, Katz and Lazarfeld [73] conducted a survey in which respondents were asked if they had recently changed their attitude on a specific topic and if so, to name who had influenced them. If the respondents did not change their attitude, it was assumed they had not been influenced. This approach of defining influence has persisted over time, leading to a framework in which social influence can mostly be a driver of change. However, as Gitlin [47] noted already in 1978 in a critique of the two-step model of communication proposed by Katz and Lazarsfeld [73]: "In the phase of high-consumption capitalism, "new" is the symbolic affirmation of positive value and "old" an emblem of backwardness thus changing one's mind about products can be considered a routine event (...). Under these conditions, one cannot take invariance for granted but as something that has to be explained".

In this article we propose a theoretical framework in which the decision to repurchase or switch to a new product is related to two types of stimuli: intrapersonal (related to the self) and interpersonal (related to the social group). To support our theoretical arguments, we conduct two randomized controlled trials and an empirical investigation in which we relate the probability to repurchase a product to variety seeking behavior [94] (intrapersonal stimuli) and exposure to information from an opinion leader or the social network (interpersonal stimuli). Our results show that variety seeking has a negative effect on the probability to repurchase, social exposure to information supporting the repurchase has a positive effect while social exposure to information supporting switching to a new product has a negative effect. This implies the effect of social influence on switching behavior (e.g. new product adoption [30, 65]) is only one facet of social influence and a potentially equally important aspect is its effect on repeated behavior.

To illustrate the implications of our theoretical framework we construct a Markov model of product choice that relates the probability to buy a product to variety seeking behavior and social exposure. The results show that when both variety seeking and social exposure have a positive valence towards product switch, the effect of social exposure on product switch is overestimated as the decision to switch is already made by the consumer and thus social exposure has merely a reinforcement effect. On the other hand, when variety seeking has a positive valence towards product switch and social exposure has a positive valence towards repurchase, a repurchase reflects an actual influence effect, as in the absence of social exposure the individual would have switched to the new product. This is to our knowledge the first study documenting such an effect and shows that individuals who do not promote opinion or behavioral change can be as important for diffusion processes as are those who promote it.

In the next section we review the literature on social influence and variety seeking behavior. Then we construct the theoretical framework that relates repurchase behavior to social exposure and variety seeking behavior. Following this, we describe the two experimental designs and the empirical study and present the results of the analysis. Next, we illustrate the implications of our theoretical framework by constructing the Markov model of product choice. We conclude with a general discussion and avenues for further research.

4.2 Literature review

Social influence as driver of behavioral change

In the seminal study of the two step model of communication flow between mass media and the public [73] Katz and Lazarsfeld conducted a survey in which respondents were asked if they had recently changed their attitude on a specific topic and if so, to name who had influenced them. If the respondents did not change their attitude, it was assumed they had not been influenced. This approach of defining influence has persisted over time. Deutsch and Gerard [34] and then Iyengar et al. [64] define informational influence as information obtained from peers that serves as evidence about reality and that changes one's beliefs about the true state of the world. Goldenberg et al. [52] asked respondents to name a person they would consult before purchasing a new product and rate the extent to which this person has influenced their decision. Van den Bulte and Joshi [135] proposed a two segment diffusion model with influential and imitators. The influentials are people who adopt a new product independently of others but who influence others (the imitators) to adopt. Trusov et al. [131] proposed a model to identify influential users in online social networks based on the effect they have on the activity level of the other users. In a randomized experiment on Facebook, Aral and Walker [11] estimated the influence of an individual on his peers by modeling the time to peer adoption as a function of the number of influence-mediating messages the peer had received. If a peer received messages from the influential and adopted shortly after, then she was influenced. Studies which did not imply change in the definition/operationalization of social influence, used change to evaluate the impact of social influence on a process of interest [30, 51, 58, 65, 13, 115]. Studies on the diffusion of innovation have shown that the opinion leaders in a community have a positive effect on the adoption of new products [30, 138, 65, 13]. In studies on public health, social leaders were used to implement new community health programs [133, 134, 77]. Overall, there is extensive research which has demonstrated the role of social influence as a driver of change in opinions, attitudes or behavior. However, often change can come from the individuals themselves, without any external trigger. Next, we discuss several studies that explain why individuals engage in varied behavior.

Variety seeking as driver of behavioral change

Extensive research on variety seeking behavior has shown that, when consumers have to choose more than one item from a choice set, they tend to vary their choices [e.g. 94], even if this leads to less preferable outcomes [114]. In their review, McAlister et al. [94] advanced two explanations of the varied behavior. The first is derived variation. The observed variety in choices can be determined by external factors that are not related to the consumer's preference for change. Examples of such factors are changes in the consumer's choice problem or different usage purposes of the product [94]. The second explanation is direct variation. Van Trijp et al. [136] call this the true variety seeking behavior and defines it as "the biased behavioral response by some decision making unit to a specific item relative to previous responses within the same behavioral category, due to the utility inherent in variation per se, independent of the instrumental or functional value of the alternatives or items". McAlister et al. [94] consider that direct varied behavior can be caused by two types of factors: intrapersonal and interpersonal. The intrapersonal motivation was linked to the existence of an ideal level of stimulation (novelty, change). When the stimulation level falls bellow the ideal level, cognitive action will produce more input (e.g. exploration, novelty seeking) [94]. That is, people will include more variety in their choices to reach back their ideal level of stimulation. This has strong implications for understanding phenomena like brand switching or new product adoption. Givon [48] has shown that brand selection is determined by: (1) the utility derived from switching the brands and (2) the underlying preferences for different brands. The interpersonal motivation states that the individual preference for varied behavior can be as well influenced by the social group. Ratner et al. [114] have shown that the amount of variety in consumption decisions is influenced by the observability of the consumption. When consumption is subjected to public scrutiny, people include more variety in their choices, as this is associated with positive personal attributes like open-mindedness. Ariely and Levav [12] have shown that in a group context people chose something other than their favorite product if that was already selected by another member, in order to assert their uniqueness.

In this article we show that social influence is a powerful determinant not only of product switching, like in the case of new product adoption, but also of product repurchase. When consumption situations induce high variety seeking behavior, social influence can act to inhibit it, resulting in higher repurchase rates. To our knowledge, the positive effect of social influence on product repurchase has only been addressed by Iyengar et al. [64]. The authors showed that consumers look for social approval of repeated consumption and thus the likelihood to repurchase is subjected to normative social influence. This implies that the decision to repurchase is already made by the consumer and will remain so as long as the social group supports it. Unlike this study, we consider the effect of social influence comes from fighting variety seeking behavior in deciding whether to repurchase or not.

4.3 Theoretical framework and hypotheses

The two streams of literature show that a purchase decision depends on two types of stimuli: intrapersonal and interpersonal. Following literature on variety seeking behavior [94], we define *intrapersonal stimuli* as the set of factors determining a repeated purchase or a product switch, that are caused by forces internal to the individual and which rely on "inherently satisfying aspects of changing behavior" [94]. Following [94], the intrapersonal stimuli are desires for the unfamiliar, alternation among familiar alternatives and need for information. The *interpersonal stimuli* are defined as the set of factors determining a repurchase or a product switch, that are caused by forces external to the individual and related to the social group. McAllister et al. [94] consider the interpersonal stimuli are the need for group affiliation and the need for personal distinction. Building on literature on social influence, we extend this definition and more generally consider as interpersonal stimuli any stimulus triggering a social influence process. The interpersonal stimuli can be either informational, when the information received from the social group serves as evidence about reality [64, 34] or normative, when behavior is determined by the desire to conform to the expectations of others [34]. The need for group affiliation and the need for personal distinction proposed in [94] can be seen as examples of normative interpersonal stimuli.

A purchase decision can be seen as one of two variations of the following unobserved two step process. In the first case (illustrated in Figure 4.1) an individual first makes the decision whether to repeat the purchase or switch to a new product driven by intrapersonal stimuli (stage I) and then confronts it with the interpersonal stimuli (stage II). In the second case, an individual first makes a decision driven by interpersonal stimuli (stage I) and then confronts it with his own, intrapersonal stimuli (stage II). For both cases, the observed purchase behavior is the result of the two processes and is determined by which type of stimuli is stronger and the order of their effects. In this article we focus of the first case, where intrapersonal stimuli act first in making the purchase decision. Extending the theoretical framework to cover the second case, thus illustrating the ordering effect, provides a promising avenue for further research.

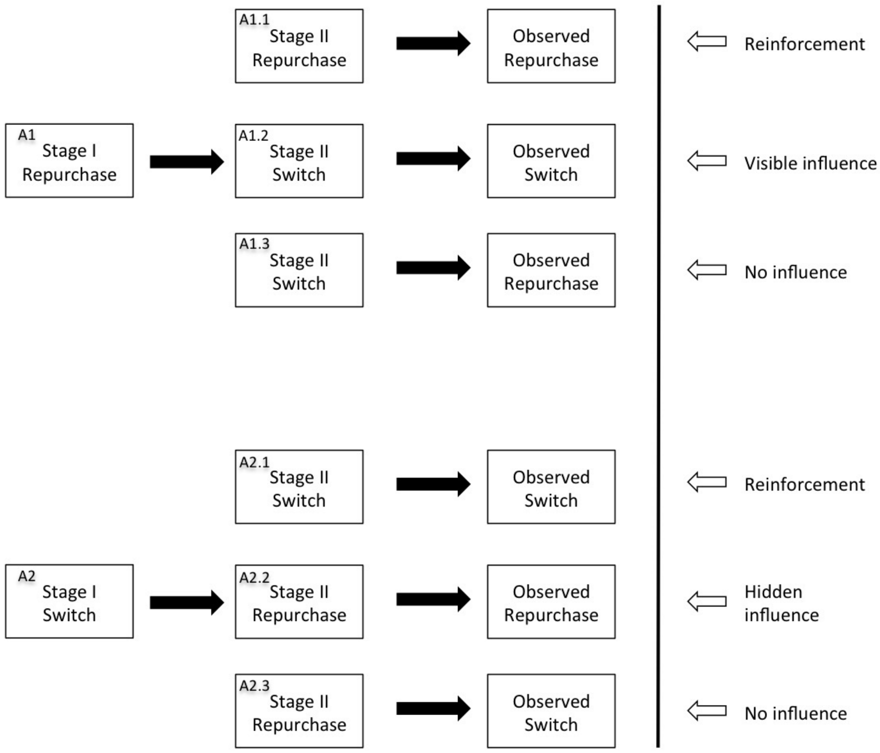


Fig. 4.1 **Theoretical framework**. Repurchase behavior is related to intrapersonal stimuli (Stage I) and interpersonal stimuli (Stage II).

If the first stage decision is to repurchase (A1) and the interpersonal stimuli have a positive valence towards it (A1.1), the social group reinforces the repurchase decision, leading to no change in behavior. If the interpersonal stimuli have a positive valence towards product switching (A1.2), the two types of stimuli are conflicting and the observed purchase behavior is determined by the stronger stimulus. If there is a product switch, the individual was influenced to switch. We call this process *visible influence* and define it as the *change* in purchase behavior that results from interaction with another individual or a group, change which would have not happened had the interaction not taken place. The process is *visible* because its effect can be directly observed as it leads to a change in purchase behavior. The above definition is similar to how influence is explicitly or implicitly defined in most social influence studies [e.g. 23, 131, 11]. The visible influence effect helps firms acquire new customers from competing brands, being thus beneficial for customer acquisition.

If the first stage decision is to switch to a new product (A2) and the interpersonal stimuli have a positive valence towards it (A2.1), the social group reinforces the switching decision, leading to a change in purchase behavior. If the interpersonal stimuli have a positive valence towards repurchase (A2.2), the observed behavior is determined by the stronger stimulus. If there is no change in the purchase behavior, the individual was influenced not to change. We call this process *hidden influence* and define it as the *lack of change* in purchase behavior that results from interaction with another individual or a group, change which would have happened had the interaction not taken place. The process is *hidden* because its effect cannot directly be observed as there is no change in purchase behavior. To our knowledge, this is the first study illustrating this effect. The hidden influence effect helps firms preserve their existing customer base, being thus beneficial to customer retention.

In the remainder of the article we focus on those cases where intrapersonal and interpersonal stimuli have conflicting effects (A1.2 & A2.2). The remaining cases can be seen as reinforcement effects of social influence on purchase behavior (A1.1 & A2.1) or as unsuccessful influence attempts (A1.3 & A2.3). Furthermore, the scope of this article is not to quantify the effect of different intrapersonal or interpersonal stimuli, as this has been extensively covered by literature on variety seeking behavior [e.g. 94] and social influence [e.g. 64, 52]. We thus make no distinction between the different intrapersonal or interpersonal stimuli and collectively refer to the interpersonal stimuli as *variety seeking* and to interpersonal stimuli as *social exposure*. Lastly, we make no distinction between the purchase of products and services and collectively refer to both as *products*. Extending the theoretical framework to cover this distinction provides a promising avenue for further research. The theoretical arguments above lead to three predictions:

Hypothesis 1 (H1): *Variety seeking has a negative effect on the probability to repurchase.*

Hypothesis 2 (H2): *Social exposure supporting product repurchase has a positive effect on the probability to repurchase.*

Hypothesis 3 (H3): *Social exposure supporting product switching has a negative effect on the probability to repurchase.*

Support for the three hypotheses would provide evidence for the underlying assumptions but does not provide direct evidence of the two stage repurchase decision process. This is not a major limitation as often theoretical mechanism are not directly observable. Rather, they are interfered from their observable consequences [e.g. 64].

4.4 Results

4.4.1 Study 1 (experimental): The restaurant choice

Previous literature has shown that brand switching behavior and new product adoption are influenced by variety seeking [e.g. 94, 136] and social exposure [e.g. 30, 65, 64]. Furthermore, the effect of variety seeking on brand switching is moderated by the product category [136]. The purpose of Study 1 is to measure the effect of variety seeking and social exposure on repurchase behavior in a product category stimulating high variety seeking behavior. We conducted the experiment using the Amazon Mechanical Turk labor market. The task required participants to imagine they won two vouchers for dinning in a Tex-Mex restaurant together with a companion. They were first asked to select a restaurant (out of four) where they would like to use the first voucher. After making the choice, participants were presented information about restaurant ranking according to a well-known food critic and then asked to select the restaurant where they would like to use the second voucher. We chose this task because existing literature has shown that food consumption induces high variety seeking behavior [137]. All four restaurants in the choice list had a comparable description and offered the same type of cuisine. Tex-Mex cuisine was selected as it is common in the United States, with typically low to medium prices, appealing to most income segments of the population. We recruited 464 participants (41% female, median age group = 35-44). The sample size was chosen to reach a power of at least 80% for the hypothesis test. The participants had to be at least 18 years old, live in the United States and were paid $0.5 for agreeing to participate.

We manipulated the restaurant ranking such that participants in the control condition were told that all four restaurants are ranked among the top ten best in town, while the participants in the treatment condition were told that the restaurant they have picked is the only one out of the four that was ranked in the top ten. The participants were randomly assigned to the treatment and control conditions, making a two cell experiment. Variety seeking behavior was operationalized as the mean of the ten items of the Exploratory Product Acquisition Scale (EPA, [18]) (five point Likert scale). In addition, we evaluated participant's preference for Tex-Mex food, measured on a five point Likert scale; gender, age and income group. Three participants were removed from analysis as they did not finish the questionnaire. To prove randomization, we ran a post-hoc analysis which shows there is no significant difference between the treatment and the control groups in terms of EPA ($p = 0.29$, Wilcoxon test); Tex-Mex preference ($p = 0.69$, Wilcoxon test); gender ($p = 0.56$, chi-square test) and income ($p = 0.47$, chi-square test). The study was pre-registered on the Open Science Framework [128].

The low repurchase rate observed in the control group (20%) confirms our choice of Tex-Mex dining as a task stimulating variety seeking behavior. In support of Hypothesis 2, the same restaurant was chosen both times by more participants in the treatment than in the control group (50% vs. 20%, $p < 10^{-10}$ Wilcoxon test). To explore this effect we run a logistic regression with the choice index as a dependent variable (1 if the same restaurant was chosen both times, 0 otherwise) and the treatment condition, variety seeking, preference for Tex-Mex food and the demographic variables as predictors. The results (see Table 4.1) show there is a significant positive effect of the treatment condition on the choice index (coefficient=1.47, SE=0.22, OR=4.36, $p < 10^{-10}$). Variety seeking has a negative, significant effect (coefficient=-0.99, SE=0.16, OR=0.37, $p < 10^{-10}$), which supports Hypothesis 1 and is in line with existing literature on variety seeking behavior [e.g. 94]. The preference for Tex-Mex (coefficient=0.07 SE=0.13, OR=1.08, p=0.56) and demographic variables were included as control variables.

So far we observed that for a product category stimulating variety seeking behavior, the repurchase decision is driven by both variety seeking and social exposure, supporting Hypotheses 1 & 2. In the next experiment we investigate if the same effects hold for a product category inhibiting variety seeking behavior.

4.4.2 Study 2 (experimental): The energy market

In Study 2 participants were required to make decisions about electricity providers. We chose this task as literature on variety seeking behavior has shown that products that are higher in involvement and less frequently purchased tend to inhibit variety seeking behavior [136] and electricity can be seen as such service. Furthermore, in several states in the US, the electricity market is deregulated. This implies that power generation companies that produce electricity cannot sell electricity directly to consumers. The electricity companies that sell electricity to end consumers buy it from the power generators and then compete with each other by offering different pricing structures, customer service benefits, and other incentives. Consumers can choose easily from a list of providers and sign up with them or cancel a contract online within minutes. Thus all consumers receive the same energy and there are no barriers in switching or renewing contracts with the providers. Furthermore, when signing up with a new provider, it can happen that new customers receive discounted prices. This has been shown to have a positive effect on brand switching behavior [e.g. 54] and affect brand choice decisions of consumers who exhibit variety seeking and reinforcement behavior [68]. Therefore, when conducting the study, we hypothesized a boundary effect of price promotions on social exposure. However, as it will be seen bellow, we did not find supporting evidence and thus the moderating role remains an interesting avenue for future research.

The experiment was conducted using the Amazon Mechanical Turk labor market and the setup was similar to Study 1. The task required participants to imagine they moved to a state with deregulated energy market and need to sign up for electricity supply with one out of four companies. They were first asked to select a company based on a short description that was provided. After making the choice, participants were presented two types of information. The first was about the ranking of the four companies according to a well-known market research agency. The second was the benefits all four companies offer to new subscribers. After receiving the information, participants were asked to imagine one year has passed in which they were satisfied with the selected company and they now have the choice to either continue the contract with the current company or switch to a different provider. We recruited 1304 participants (52% female, median age group = 25-34). The sample size was chosen to reach a power of at least 80% for the hypothesis test. The participants had to be at least 18 years old, be located in the United States and were paid \$0.5 for agreeing to participate. The randomization occurred along two dimensions: company ranking (2 levels) and discount offered (3 levels), resulting in a 2x3 full factorial design. Participants in the company control condition were told that all four companies are ranked among the top ten best. Participants in the company treatment condition were told that the provider they picked is the only one out of the four that was ranked in the top ten. For each of the two company ranking conditions we created three more conditions by changing the size of the discount offered: 25% off the first month's electricity bill, 50% off the first month's electricity bill, first month electricity bill free. Similarly to the first study, we evaluated participants' variety seeking behavior, measured as the mean of the ten items of the Exploratory Product Acquisition Scale [18](five point Likert scale); their gender, age and income group. In addition to the first study, we also evaluated the attitude of the participants towards taking risks as this has been found to be a major motivator for purchasing new products [113]. The attitude towards taking risks was measured as the mean of the seven items of the Instrumental SIRI Scale [141] (five point Likert scale). Thirty participants were removed from analysis, as they did not finish the questionnaire. To prove randomization, we ran a post-hoc analysis which shows there is no significant difference between the treatment and the control groups in terms of EPA ($p < 0.11$, anova test); SIRI ($p < 0.38$, anova test); gender ($p < 0.68$, chi-square test) and income ($p < 0.47$, chi-square test). The study was pre-registered on the Open Science Framework [129].

	Study 1				Study 2			
	Coefficient	Std.E.	OR	z-value	Coefficient	Std.E.	OR	z-value
Intercept	1.11	0.67	3.04	0.10	1.99	0.43	7.32	$< 10^{-6***}$
OL Treatment	1.47	0.22	4.36	$< 10^{-10***}$	1.43	0.13	4.17	$< 10^{-10***}$
Discount	-	-	-	-	- 0.005	0.21	0.99	0.98
Variety seeking	-0.99	0.16	0.37	$< 10^{-10***}$	-0.74	0.10	0.88	$< 10^{-10***}$
Risk attitude	-	-	-	-	-0.75	0.10	0.93	0.45
Tex Mex pref.	0.07	0.13	1.08	0.56	-	-	-	-
Gender$^{(1)}$	0.41	0.23	1.51	0.07*	-0.18	0.13	0.83	0.17
Age$^{(1)}$	0.13	0.1	1.14	0.19	0.09	0.05	1.09	0.08*
Income$^{(1)}$	-0.06	0.09	0.94	0.50	-0.01	0.05	0.99	0.98

Table 4.1 **Results of controlled studies**. $^{(1)}$ Coefficient values are based on a different model where $8/32$ observations have been deleted due to missingness. Age and Income are considered continuous.

The higher repurchase rate observed in the company control condition compared to Study 1 (50% vs 20%) confirms that choice of energy providers inhibits variety seeking behavior. We run a logistic regression with the choice index as a dependent variable (1 if the same electricity provider was chosen both times, 0 otherwise) and the company ranking, variety seeking, discount offered, risk attitude and the demographic variables as predictors. The results (see Table 4.1) show that variety seeking has a negative, significant effect (coefficient=-0.74, SE=0.10, OR=0.49, $p < 10^{-10}$), bringing further support to Hypothesis 1. The company ranking condition has a positive, significant effect (coefficient=1.41, SE=0.13 OR=4.1 $p < 10^{-10}$), bringing further support to Hypothesis 2. In line with existing research, the risk attitude (coefficient=-0.13, SE=0.10, OR=0.88, $p < 0.19$) has a negative, however not significant effect. The demographic variables are included as control variables. Interestingly, we did not find any significant effect of the discount offered on the probability to repurchase. We see two possible explanations: (1) the choice of discount levels was not well calibrated and the treatments were not perceived as different enough; (2) the effect of social exposure is robust to small increases in the discounts offered. As the second explanation was not hypothesized before conduction the study, further research will need to be conducted to asses if social exposure is indeed robust to small changes or if the non-significant effects are an artifact of the design.

4.4.3 Study 3 (empirical): The online recipe community

Data description. Studies 1 & 2 have shown that variety seeking and social exposure supporting the repurchase have a positive effect on the probability to repurchase. What still

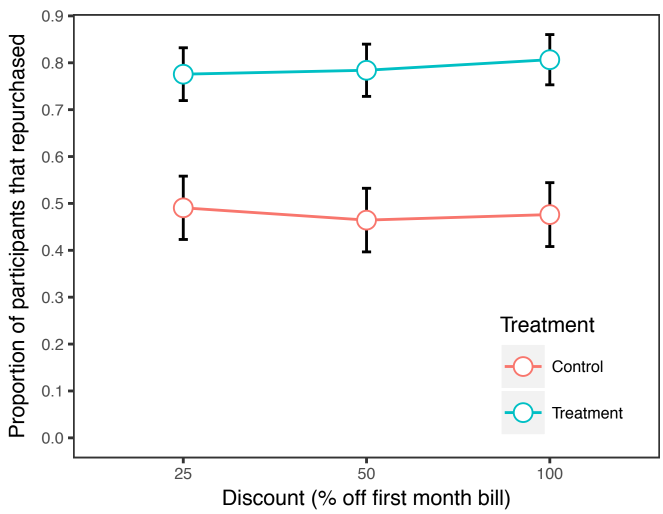


Fig. 4.2 **Study 2: The effect of social exposure on repurchase rate**. The x axis represents the discount offered to new customers. The y axis represents the repurchase rate. The color is given by the company treatment.

remains to be shown is that social exposure supporting product switch has a negative effect on the probability to repurchase (Hypothesis 3). Furthermore, while experimental studies are a good approach to asses the effects in isolation, by eliminating the potential for selection bias, the applicability of the results outside the experimental environment is often criticized [57]. In consequence, an empirical study provides a strong contribution to assessing the validity of the effects. The purpose of Study 3 is to evaluate the effect of social exposure supporting product switch and of social exposure supporting product repurchase on the probability to repurchase in a non-experimental setup. We conducted our empirical analysis in the context of food choice as in Study 1. Our data comes from a manufacturer of kitchen appliances that hosts an online community where members share their cooking experience. Registered users can view, create and rate recipes, discuss through private messages with other members and establish friendship connections. A member can see all recipes created by any member, whether they are friends or not. Each action taken on the platform is rewarded with a number of points, which reflects the engagement of the user in the community. We were given complete access to records containing all actions taken on the platform by the community members between June 2011 and November 2015. The dataset consists of six separate communities, each from a different country across two continents. Overall, the dataset contains 1,383,107 registered members creating, sharing, commenting and rating 103,373 recipes within the four years. A detailed description can be found in Table B.1 in Appendix B.1. The dataset includes three key pieces of information that are relevant for our analysis: recipe data, network data and activity information.

Recipe data. The main value of the community relies in the collection of recipes that are created by its members. Every recipe is part of one of 10-16 recipe categories which define the meal course. Examples include: soups, starters, main dish fish, main dish meat, desert, etc. The categories are decided by the platform managers and there is a slight difference in the number of categories between the six communities. A complete list of categories in each community together with descriptive statistics can be found in Appendix B.1. From the total number of 1,383,107 users, only 328,910 are active, where activity is defined as making at least one action on the platform except registering and logging in. From the active users, 33,653 create recipes with an average number of 3.072 recipes per user who creates recipes (sd=9.25, median=1, max=739). On average, an user creates recipes in 1.87 categories (sd=1.58, median=1, max=16), with an average of 1.3 recipes per category (sd=0.90, median=1, max=46.18). The average member tenure (lifetime on the platform) for users creating recipes is 2.32 years (sd=1.28, min=0.002, median=2.19 max=4.45) with a mean tenure during which a user is actively creating recipes of 0.66 years (sd=0.66, median=0.31, max=4.4). In addition to creating recipes, community members can evaluate the available recipes through a 5-star rating system (1=lowest, 5=highest). Over the four years, the users performed 560,068 ratings. During this time, 71,128 users made at least two ratings, with an average number of 6.85 ratings per user who made at least two ratings (sd=15.25, median=3 max=1090).

Network data. The community has as well a social function. Users can create friendship connections and communicate with each other through private messages. The resulting friendship network is undirected (two users can be friends or not) and unweighted. However, a tie between two nodes (users) in a network can be weighted by the number of private messages sent. Even though the messages have a sender and a receiver (and thus a sense of directionality), similarly to [104, 115] we consider the resulting weighted network as undirected. During the four years, 67,556 users made at least one friendship connection, 36,497 users sent at least one message, 84,839 received at least one message and 29,887 users have friends who created at least one recipe.

Activity data. From the activity of the users we can derive their engagement in the community, measured as the number of user points earned and their tenure, measured as the time elapsed since they joined the community. The data has a time resolution of one second and all actions taken on the platform have a corresponding time-stamp.

The Appendix contains distribution plots for all the above variables within each community (Figures B.1 - B.10) and in the aggregated data (Figure B.11), together with pairwise relationships between them (Figure B.12).

Methodology. We consider that rated recipes have been previously experienced and use this as a proxy for consumption. Thus even though rating a recipe is different from an actual consumption situation, like dinning in a restaurant, we consider the findings can be extended to product purchase as both situations involve a cognitive effort (deciding on a recipe/restaurant), financial effort (buy ingredients/pay the meal), time investment (cooking time/traveling time) and both result in a certain level of satisfaction (enjoy or not the recipe cooked / meal ordered). There is substantial literature showing that food consumption is characterized by variety seeking behavior [e.g. 137]. We assume that in the absence of any confounding factors, users do not always try recipes from the same category but explore alternative options. We analyze the recipe ratings of the community members and investigate how often the same category is repeated in two consecutive ratings. We use r to define the current rating, $r-1$ to define the previous rating, $C(i,m)$ to define the category corresponding to recipe rated by user i in rating m and $T(m)$ to define the timestamp corresponding to rating m. We will further use the term focal category to define the category corresponding to the recipe rated in the previous rating (C(i, $r-1$)). We employ a mixed effects logistic regression in which the dependent variable (DV), repeat rate, is the repetition of recipe category:

$$DV = \delta(C(i,r),C(i,r-1)) \tag{4.1}$$

where $\delta(A,B)$ is a binary variable taking value one if at least one element of A is in B and 0 otherwise. In consequence, the dependent variable takes value 1 if the recipe rated at time t belongs to the focal category and 0 otherwise. As predictors we use the variables described bellow.

Exposure hidden. Following [104, 115], the exposure variable represents the presence of neighbors in the social network of the focal individual who support product repurchase or product switch. Thus we measure exposure as the number of friends (direct neighbors in the social network) who created at least one recipe in the focal category up to rating time $T(r)$. More specifically, the exposure of customer i at time $T(r)$ to friends supporting repurchase is defined as:

$$\text{Exposure hidden}_{i,T(r)} = \sum_{j \in A_{i,T(r)}} \delta(C(j,m)_{T(m)<T(r)},C(i,r-1)) \tag{4.2}$$

where $A_{i,T(r)}$ represents the direct neighbors of i at rating time $T(r)$. Similar to [104], in addition to the basic exposure we consider as well the lagged exposure. This is defined as the number of friends who created at least one recipe in the focal category between the previous and the current rating:

$$\text{Exposure hidden lagged}_{i,T(r)} = \sum_{j \in A_{i,T(r)}} \delta(C(j,m)_{T(r-1) \leq T(m) < T(r)}, C(i,r-1)) \tag{4.3}$$

Exposure visible. Represents the number of friends who created at least one recipe that is not in the focal category up to rating time $T(r)$:

$$\text{Exposure visible}_{i,T(r)} = \sum_{j \in A_{i,T(r)}} \bar{\delta}(C(j,m)_{T(m) < T(r)}, C(i,r-1)) \tag{4.4}$$

where $\bar{\delta}(A,B)$ is a binary variable taking value 1 if at least one element of A is not in B and 0 otherwise. We consider as well the lagged exposure, defined as the number of friends who created at least one recipe not in the focal category between the previous and the current rating:

$$\text{Exposure visible lagged}_{i,T(r)} = \sum_{j \in A_{i,T(r)}} \bar{\delta}(C(j,m)_{T(r-1) \leq T(m) < T(r)}, C(i,r-1)) \tag{4.5}$$

Tie strength hidden. Consistent with previous studies [104, 115], we use the volume of communication between two users as an indicator of tie strength. We define the communication volume between user i and j at time $T(m)$ ($COM(i,j,T(m))$) as the sum of messages exchanged between i and j (sent or received) divided by the total number of messages i exchanged up to time $T(m)$. Similarly to [104, 115], we compute the tie strength as the average communication volume with friends that created at least one recipe in the focal category:

$$\text{Tie strength hidden}_{i,T(r)} = \frac{\sum_{j \in A_{i,T(r)}} COM(i,j,T(r)) \delta(C(j,m)_{T(m) < T(r)}, C(i,r-1))}{\sum_{j \in A_{i,T(r)}} \delta(C(j,m)_{T(m) < T(r)}, C(i,r-1))} \tag{4.6}$$

If there are no friends who created at least one recipe in the focal category then Tie strength hidden$_{i,T(r)}$ = 0. We consider as well the lagged tie strength, defined as the average communication volume with friends who created at least one recipe in the focal category between the previous and the current rating.

Tie strength visible. Similarly as above, it is computed as the average communication volume with friends that created at least one recipe that is not in the focal category while lagged tie strength is the average communication volume with friends who created between the previous and the current rating at least one recipe that is not in the focal category.

$$\text{Tie strength visible}_{i,T(r)} = \frac{\sum_{j \in A_{i,T(r)}} COM(i,j,T(r))\bar{\delta}(C(j,m)_{T(m)<T(r)}, C(i,r-1))}{\sum_{j \in A_{i,T(r)}} \bar{\delta}(C(j,m)_{T(m)<T(r)}, C(i,r-1))} \quad (4.7)$$

Homophily hidden. Homophily measures how similar individuals are in terms of defined characteristics. We consider there is a high homophily between two users if they are interested in the same recipes. More formally, we compute the pairwise homophily between two users i and j at $T(r)$ ($H(i,j,T(r))$) as the Jaccard similarity (i.e. normalized set overlap) between the set of recipes the users rated up to time $T(r)$. If both users rated exactly the same recipes, the pairwise homophily is one. On the other hand, if the two users rated completely different recipes, the pairwise homophily is zero. The homophily score is then computed as the average pairwise homophily over the friends who created at least one recipe in the focal category up to time $T(r)$:

$$\text{Homophily hidden}_{i,T(r)} = \frac{\sum_{j \in A_{i,T(r)}} H(i,j,T(r))\delta(C(j,m)_{T(m)<T(r)}, C(i,r-1))}{\sum_{j \in A_{i,T(r)}} \delta(C(j,m)_{T(m)<T(r)}, C(i,r-1))} \quad (4.8)$$

We include as well the lagged homophily, defined as the average pairwise homophily over the friends who created at least one recipe in the focal category between the previous and the current rating.

Homophily visible. Similarly as above, we compute the homophily score as the average pairwise homophily over the friends who created up to time $T(r)$ at least one recipe that is not in the focal category and the lagged homophily as the average pairwise homophily over the friends who created between the previous and the current rating at least one recipe that is not in the focal category:

$$\text{Homophily visible}_{i,T(r)} = \frac{\sum_{j \in A_{i,T(r)}} H(i,j,T(r))\bar{\delta}(C(j,m)_{T(m)<T(r)}, C(i,r-1))}{\sum_{j \in A_{i,T(r)}} \bar{\delta}(C(j,m)_{T(m)<T(r)}, C(i,r-1))} \quad (4.9)$$

Satisfaction. An important driver of category change can be dissatisfaction with previous choice. We measure satisfaction as the rating value of last recipe rated. As controls we consider the number of friends at time $T(r)$ (*Degree*); the number of recipes in the focal category at time $T(r)$ (*Category size*); number of recipes in all categories except the focal category at time $T(r)$ (*Available alternatives*); *Inter-event time* (in days) between two consecutive ratings $(T(r) - T(r-1))$; *Mean inter-event time* defined as the mean inter-event time between two ratings and measured in days ($\sum_{m \leq r}(T(m) - T(m-1))/r$); number of messages sent by the user up to time $T(r)$ (*Communication volume*); nr of recipes rated by the user until $T(r)$ (*Recipes rated*); number of recipes created by the user until $T(r)$ (*Recipes created*), number of user points earned up to time $T(r)$ (*Engagement*); number of user points earned in the last period (*Recent engagement*); time (in years) elapsed since registration or the start of the observation period (*Tenure*) and community membership (*Community*). To account for any other source of heterogeneity between users we introduce a random intercept.

Results and discussion. The data used for analysis includes information from 71,128 unique users in 6 communities who performed 416,633 recipe ratings. As we are interested in the repetition of a recipe category, all users who did not rate at least two recipes have been excluded from the analysis (257,782 active users). The unit of analysis is the recipe rating.

Before running the formal model, consider the following descriptive analysis. Out of the 416,633 observations (recipe ratings), in 289,481 the users do not have any friends in the social network (the number of friends is cumulative, thus a user can have no friends for early observations and have friends for late observations). This allows estimating the rate of repeating the category in the absence of social influence. For observations where individuals do have friends (N=127,152 ratings) three interesting cases are: there is only exposure to hidden influence (N=1,689), there is only exposure to visible influence (N=20,810) and there is exposure to both(N= 52,810). In the first case, the social influence effect, if any, can only be attributed to hidden influence, in the second to visible influence, while in the third to both. Figure 4.3 shows the results. It can be seen that in the absence of social influence, the mean repeat rate is 20% (N=289,481, se= 0.0007). Having only exposure to friends who support repeating the category (hidden influence, blue color) increases the repeat rate to 30% (N=1,689, se=0.011), having exposure only to friends supporting changing the category (visible influence, red color) decreases the repeat rate to 17% (N= 20,810, se=0.002), while having exposure to both results in a repeat rate of 23% (N= 52,810, se=0.002). Thus exposure to friends who support repeating the category seems to increase the repeat rate, exposure to friends who support changing the category decrease it while their combined effect lies in-between the two extremes.

To investigate this effect formally, we restrict the analysis to observations where individuals have at least one friend at the time of the observation (127,152 ratings; 15,168 unique users). Table 4.2 shows the results of the mixed effects logistic model, while Figure B.13 in Appendix B.2 shows the distribution of the exposure variables. Exposure hidden, Exposure hidden lagged, Tie strength hidden, Tie strength hidden lagged, Homophily hidden and Homophily hidden lagged have a positive, significant effect. This shows that the more friends who have recipes in the focal category, the higher the likelihood to repeat the category,

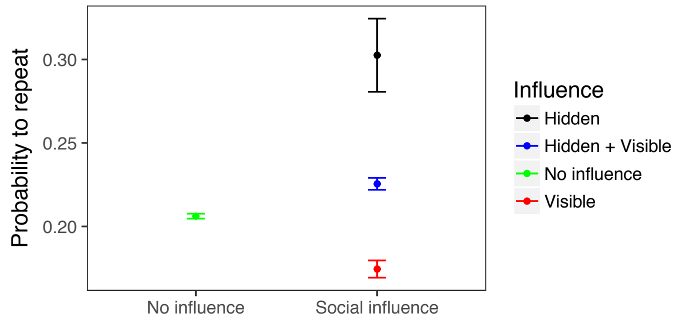


Fig. 4.3 **Effect of hidden and visible influence (Descriptive analysis)**

bringing supporting evidence to H2. Exposure visible, Exposure visible lagged, Tie strength visible, Tie strength visible lagged, Homophily visible and Homophily visible lagged have a negative, significant effect. This shows that the more friends who create recipes in non-focal categories, the lower the likelihood to repeat the category, bringing supporting evidence to H3. The non-social variables (highlighted in grey) have the expected sign of the effect. The size of the focal category has a positive significant effect, the number of alternative recipes a negative significant effect, the inter-event time and mean inter-event time a negative, significant effect. The only exception is satisfaction which has a non-significant effect. One explanation for the non significant effect of satisfaction is that the variable is very skewed towards high ratings (92% of the observation have a rating of at least 4 out of 5; see Figure B.14 in Appendix B.2) and thus there is not enough variation in the data.

Alternative model specifications When constructing the model above we made two choices that could influence the results: (1) the time-lag (number of previous ratings) used to compute the lagged variables and (2) the operationalization of social exposure. First, to check the robustness of the results to the selection of time-lag, we re-run the model using a lag of three, five and seven previous ratings. For lag three there was no change in the sign or significance of the exposure variables. The same results hold for lags five and seven, with the exception of Exposure visible. The coefficient has the same sign but it is not significant. The reasons is that it's effect is captured by Exposure visible lagged, as Exposure visible and Exposure visible lagged become highly correlated (Pearson r = 0.77 for lag seven; 0.72 for lag five; 0.65 for lag three; 0.47 for lag one).

Second, to show the results are robust to how we operationalize social exposure we consider two alternative specifications. The first approach is to normalize the number of supporting friends by the size of the neighborhood (model M1). That is, we define social exposure as the number of friends having at least one recipe in / not in the focal category up

	Coefficient	SE	p-value
(Intercept)	-1.50	6.05×10^{-2}	$< 10^{-10***}$
Exposure hidden	3.73×10^{-2}	4.12×10^{-3}	$< 10^{-10***}$
Exposure hidden lagged	2.45×10^{-1}	2.28×10^{-2}	$< 10^{-10***}$
Exposure visible	-1.99×10^{-1}	4.35×10^{-3}	$< 10^{-6***}$
Exposure visible lagged	-1.43×10^{-1}	1.47×10^{-2}	$< 10^{-10***}$
Tie strength hidden	2.46×10^{-1}	5.60×10^{-2}	$< 10^{-5***}$
Tie strength hidden lagged	5.74×10^{-1}	1.25×10^{-1}	$< 10^{-6***}$
Tie strength visible	-1.48×10^{-1}	4.89×10^{-2}	$2.22 \times 10^{-3**}$
Tie strength visible lagged	-5.22×10^{-1}	9.26×10^{-2}	$< 10^{-8***}$
Homophily hidden	2.29	5.67×10^{-1}	$< 10^{-5***}$
Homophily hidden lagged	3.69	9.52×10^{-1}	$< 10^{-4***}$
Homophily visible	-1.75	5.73×10^{-1}	$5.37 \times 10^{-2.}$
Homophily hidden lagged	-2.65	7.75×10^{-1}	$6.33 \times 10^{-4***}$
Degree	-1.80×10^{-3}	1.84×10^{-3}	5.21×10^{-1}
Satisfaction	-1.11×10^{-2}	1.05×10^{-2}	2.91×10^{-1}
Category size	2.88×10^{-4}	4.09×10^{-6}	$< 10^{-10***}$
Available alternatives	-3.70×10^{-5}	1.43×10^{-6}	$< 10^{-10***}$
Inter-event time	-1.05×10^{-3}	1.05×10^{-4}	$< 10^{-10***}$
Mean inter-event time	-1.40×10^{-4}	6.23×10^{-5}	$2.52 \times 10^{-2*}$
Recipes rated	1.52×10^{-4}	1.67×10^{-4}	3.69×10^{-1}
Recipes created	2.54×10^{-3}	4.46×10^{-4}	$< 10^{-8***}$
Communication volume	-3.96×10^{-6}	1.75×10^{-6}	$2.36 \times 10^{-2*}$
Engagement	-4.35×10^{-6}	1.01×10^{-5}	6.67×10^{-1}
Recent engagement	-1.83×10^{-4}	1.97×10^{-4}	3.53×10^{-1}
Tenure	-2.29×10^{-4}	1.90×10^{-2}	9.90×10^{-1}
Community I	1.02×10^{-1}	3.86×10^{-2}	$7.90 \times 10^{-3**}$
Community II	2.61×10^{-1}	4.50×10^{-2}	$< 10^{-9***}$
Community III	-9.22×10^{-2}	4.23×10^{-2}	$2.95 \times 10^{-2*}$
Community IV	5.36×10^{-1}	7.04×10^{-2}	$< 10^{-10***}$
Community VI	7.19×10^{-1}	6.88×10^{-2}	$< 10^{-10***}$

Table 4.2 Results Study 3.

to time $T(r)$ divided by the total number of friends at time $T(r)$, with a similar definition for the lagged variables. A different approach that does not directly rely on the number of supporting friends, is to define exposure in terms of the number of recipes created by friends (model M2). Following this, Exposure hidden can be defined as the number of recipes in the focal category created by friends up to time $T(r)$, while Exposure visible can be defined as the number of recipes created by friends up to time $T(r)$ that are not in the focal category. Similarly as above, the lagged variables are computed by taking into account only the recipes created between the previous and the current rating. We re-run the analysis using both specifications (based on the average number of supporting friends and number of recipes). The results show no change in the sign and significance of the exposure variables (see Tables B.8 and B.9 in the appendix).

Study 1 and Study 2 brought supporting evidence for H1 and H2 and Study 3 brought evidence for H2 and H3. This illustrates how variety seeking and social exposure relate to repurchase behavior. In the next step we investigate how the effects hypothesized relate to the overall probability to buy a product, either as first purchase or as repurchase.

4.5 Implications: The impact of variety seeking and social exposure on product choice

To illustrate the differential impact of variety seeking and social exposure on the probability to buy a product, we consider the following Markov model of consumer choice. A similar model was used by Kahn [68] to illustrate the effect of price promotions on variety seeking. In a repeated purchase context, we consider the product choice at time t depends on three factors: the previous purchase, variety seeking [94, 136] and social exposure [e.g. 65]. As mentioned above, we distinguish between social exposure supporting product switch (exposure visible) and social exposure supporting repurchase (exposure hidden). To ease the explanation we consider a market consisting of two competing products from the same product category (Product 0 and Product 1), with an opinion leader advocating one of the products (Product 1). We assume variety seeking is induced only by product category [136], with all consumers having the same level of variety seeking within a given category. Furthermore, the company acts in a market with no competitors employing similar strategies.

When none of the products is advocated by the opinion leader, the choice behavior is described by the Markov matrix in Table 4.3. A customer purchases a new product with probability v, while repurchases the currently owned product with probability (1-v). The

t=0 / t=1	Product 1	Product 0
Product 1	$1-v$	v
Product 0	v	$1-v$

Table 4.3 Product choice with no social influence

t=0 / t=1	Product 1	Product 0
Product 1	$(1-v)+E_h v$	$v-E_h v$
Product 0	$v+E_v(1-v)$	$(1-v)-E_v(1-v)$

Table 4.4 Product choice with social influence

probability v is determined by variety seeking and takes values close to 1 when individuals engage in varied behavior.

When Product 1 is advocated by the opinion leader, the choice behavior is described by the Markov matrix in Table 4.4. The probability to buy the advocated product depends on variety seeking (v) and the two social exposure variables: exposure hidden (E_h) and exposure visible (E_v). If at $t=0$ Product 1 was purchased, the probability to repurchase Product 1 at $t=1$ is computed as the probability to decide to repurchase driven by variety seeking $(1-v)$ plus the probability to decide not to repurchase (v) multiplied by the probability to be influenced to repurchase the advocated product (E_h). On the other hand, if at $t=0$ Product 0 was purchased, the probability to buy Product 1 at $t=1$ is computed as the probability not to repurchase, driven by variety seeking (v), plus the probability to decide to repurchase $(1-v)$ multiplied by the probability to be influenced to buy the new (advocated) product (E_v). This model implies that the effect of both hidden and visible exposure on purchase behavior depends linearly on variety seeking. Furthermore, variety seeking governs which type of exposure is stronger. Increasing variety seeking decreases the effect of exposure visible and increases the effect of exposure hidden. Based on the switching matrix in Table 4.4, we can derive the stationary distribution of the Markov chain and computed the long run probability of buying Product 1 (P(1)). The long run probability can be interpreted as the market share of Product 1 in the limit of large time. It is defined as:

$$P(1) = \frac{1-(1-v)(1-E_v)}{2-(1-v)(2-E_h-E_v)-E_h} \tag{4.10}$$

Based on the long run probability to buy Product 1, we can derive qualitative insights about the effect of hidden and visible exposure and variety seeking on the probability to buy the advocated product. To this end we investigate how P(1) changes by increasing either E_h

or E_v, keeping everything else constant. Van Trijp [136] has shown that the effect of varied behavior on choice variety is moderated by product-level characteristics. Products that are lower in involvement, more frequently purchased, less preference-driven, and characterized by small perceived differences among brands tend to stimulate variety seeking [136]. To illustrate the effect of visible and hidden exposure on product choice we consider two product categories: the first stimulating variety seeking behavior (here v=0.9, e.g. beer [136]) and the second inhibiting variety seeking behavior (here v=0.1, e.g. cigarettes [136]). Figure 4.4 shows that for products stimulating high variety seeking behavior, exposure visible has a small, positive, increasing effect while exposure hidden has a large, positive, increasing effect. On the other hand, for products inhibiting variety seeking behavior, exposure visible has a large positive, decreasing effect while exposure hidden has a smaller, positive, increasing effect. Furthermore, the effect of exposure visible is bounded by variety seeking, as there is always a fraction v of customers who switch because of variety seeking. After this point (point B in the left panel of Figure 4.4), the effect of exposure hidden becomes stronger than the effect of exposure visible. The above qualitative findings describe the effect of one exposure variable when the other has a small, non-zero effect. An interesting case is when the effect of exposure visible is exactly zero. In this case, the model implies that the effect of exposure hidden is the same for all levels of variety seeking (see Figure B.15 in Appendix B.3). Thus the black dashed and solid lines in both panels of Figure 4.4 would overlap and thus will be no crossing at point A in the left panel of Figure 4.4.

The practical relevance of the model implications is summarized in Table 4.5 and can be described as follows. For products stimulating a high need for variety, variety seeking is the main driver of change. In such markets, in the absence of social (*hidden*) influence, consumers will frequently switch products to satisfy their variety seeking behavior. This implies that social marketing campaigns targeted to new customers (thus based on visible influence) have little potential effect as consumer will likely try the new product anyway. On the other hand, marketing campaigns targeted at existing customers could be very effective as through hidden influence consumers can be determined to go against their variety seeking behavior and repeatedly consume the same product. For products inhibiting variety seeking, the main driver of change is visible influence and the main driver of the resistance to change is low variety seeking. In such markets, in the absence of social (*visible*) influence, consumers will rarely change products and thus social marketing campaigns targeted to new customers can be very effective.

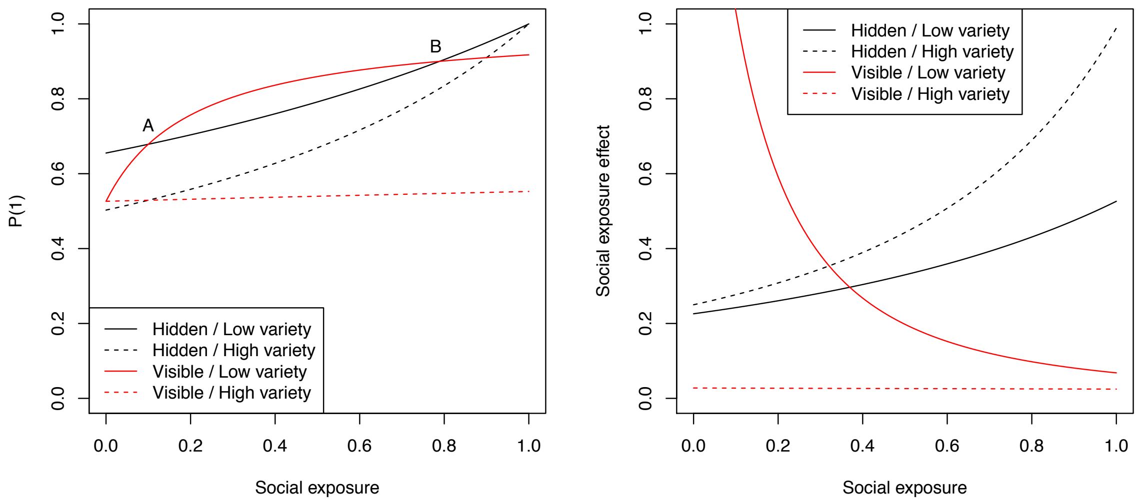


Fig. 4.4 **Effect of social exposure on the probability to purchase**. The x axis represents social exposure. It can be interpreted as the probability that the second stage decision is to repurchase (hidden) or to switch (visible). **Left panel: The effect of social exposure on the probability to buy the advocated product**. The y axis represents the probability to buy the advocated product. **Right panel: The effect of social exposure on the change in the probability to buy the advocated product**. The y axis represents the social exposure effect (the value of the derivative of P(1) with respect to either E_h or E_v).

Variety seeking	Social exposure	
	Hidden	Visible
High	+++	+
Low	+	+++

Table 4.5 **Effect of social exposure and variety seeking on product choice**. For products stimulating high variety seeking behavior (upper row) exposure hidden has a higher potential effect. For products inhibiting variety seeking behavior, exposure visible has a higher potential effect.

4.6 Discussion

In this article, we constructed a theoretical framework that relates product repurchase to variety seeking and social exposure. Study 1 showed that for products stimulating variety seeking behavior, variety seeking has a negative effect on the probability to repurchase and social exposure having a positive valence towards product repurchase has a positive effect. Study 2 showed the same effects hold for product categories inhibiting variety seeking behavior. Study 3 replicated the positive effect of social exposure having a positive valence towards product repurchase and revealed that social exposure having a positive valence towards product switching has a negative effect on the probability to repurchase. Our results contribute to existing literature on social influence and social contagion by bringing a fresh perspective on understanding the effects of social influence. Most extant studies focused on explaining *change* in purchase behavior as a consequence of social influence. However, this is only the *visible* side of social influence. For purchase occasions characterized by high variety seeking behavior, consumers typically engage in varied behavior even in the absence of any interpersonal stimuli. As our results show, in such situations consumers can be influenced to fight their variety seeking behavior and remain product loyal. How do these findings relate to the overall probability to buy a product? The results of the theoretical Markov model of product choice show that for products stimulating high variety seeking behavior, consumers are likely to try new products in the absence of any external stimuli. For companies offering such products, there is a bigger challenge in preserving the existing customers than in acquiring new customers. Thus influencer marketing campaigns targeted to existing customers have a higher potential effect. For products inhibiting variety seeking behavior, consumers will rarely switch to new products just to satisfy their need for variety. For companies offering such products, it is more challenging to acquire new customers than to preserve existing ones. Thus influencer marketing campaigns targeted at new customers have a higher potential effect.

The theoretical framework developed in this article consists of a two stage repurchase decision process, in which the first stage decision is driven by intrapersonal stimuli and the second stage decision by interpersonal stimuli. However, it is easy to imagine consumption situations where the reverse is also true. Extending the theoretical framework to cover such cases provides a promising avenue for future research. Furthermore, the implications we derived highlight how customers react to strategies of one company when no competitor is employing a similar campaign. An investigation of the effects of social influence in a market with several competitors employing the same strategy can provide an important contribution. We note that when designing Study 1 and Study 2, we initially considered variety seeking as control variable. Later, as the story developed, we decided to present the effect of variety seeking as a hypothesis in order to have a complete theoretical framework. This raises no question of post-hoc hypothesizing the effects [76] as, even though not explicitly stated as a hypothesis, the effect of variety seeking on product repurchase was included in the design (and pre-registered) before conducting both studies [128, 129].

Chapter 5

Controlling complex policy problems

Notwithstanding the usefulness of system dynamics in analyzing complex policy problems, policy design is far from straightforward and in many instances trial-and-error driven. To address this challenge, we propose to combine system dynamics with network controllability, an emerging field in network science, to facilitate the detection of effective leverage points in system dynamics models and thus to support the design of influential policies. We illustrate our approach by analyzing a classic system dynamics model: the World Dynamics model. We show that it is enough to control only 53% of the variables to steer the entire system to an arbitrary final state. We further rank all variables according to their importance in controlling the system and we validate our approach by showing that high ranked variables have a significantly larger impact on the system behavior compared to low ranked variables [1].

Keywords: network controllability | system dynamics | control centrality | driver nodes

[1]**Author Statement:** This work is in collaboration with Lukas Klaus Schoenenberger from the University of Bern, Switzerland. The article is published as Schoenenberger, L., and Tanase, R. (2017). Controlling complex policy problems: A multimethodological approach using system dynamics and network controllability. *Journal of Simulation*,1-9, doi: 10.1080/17477778.2017.1387335. The work of Radu Tanase was supported by the University Research Priority Program “Social Networks” of the University of Zurich.

5.1 Introduction

System dynamics (SD), an approach to modeling and simulating complex systems, has repeatedly demonstrated its value in contributing to the understanding and solution of complex policy problems—most notably in areas such as public health, energy and the environment, social welfare, sustainable development and security [46, 126]. Particularly in large (complex) SD models, however, the detection of model levers, i.e., variables capable of effectively and efficiently controlling complex policy problems, is a challenge. This is due to the high degree of interdependent model variables and nonlinear relationships typically present within these models. So, notwithstanding the usefulness of SD in the analysis of complex policy problems, the solution identification process (policy design) is far from trivial and in most cases trial-and-error driven [41, 107]. To address this challenge, we propose a multimethodological approach combining SD with network controllability to enhance the speed and quality of model lever discovery in SD models. In this respect, our article is a first attempt to bring SD a step closer to a very recent and fast growing field with strong roots in complex systems research: network science; thereby abiding by Anderson's [9] and Barlas' [16] call to reach out and partner with emerging systemic disciplines.

In their first editorial in Network Science, i.e., a novel journal published by Cambridge University Press, Brandes et al. [22] define the field 'as the study of the collection, management, analysis, interpretation, and presentation of relational data.' In essence, a network is a collection of points, i.e., vertices or nodes, joined together in pairs by lines, i.e., edges [103]. Networks are ubiquitous, ranging from neural networks capturing the connections between the neurons in the brain, to social networks mapping human interactions or trade networks representing the exchange of goods and services. Networks are at the heart of complex systems and consequently a deep understanding of the former has to be developed to fully understand the latter [14].

Within network science a powerful stream of research has emerged that deals with network controllability [87]. This represents the ability to steer a dynamical system from any initial state to any desired final state within finite time using suitable inputs. The methodology builds on nonlinear dynamics and control theory [87, 86, 70, 69]. In an Nature article, Liu et al. [87] presented analytical tools to identify the minimum number of driver nodes N_D in an arbitrary complex directed network that, if appropriately manipulated, can offer full control over the network. Interestingly, these analytical tools are grounded on the assumption that the controllability of nonlinear systems (networks) is often structurally similar to and determined by the system's linearized dynamics [45, 123]. In other words, for the detection of the minimum number of driver nodes N_D in nonlinear dynamic networks, network scientists revert to the same method system dynamicists have been using in eigenvalue elasticity analysis (EEA) for a long time: approximate nonlinear dynamic systems with linearized systems near their equilibrium points [107, 106].

In this paper, we show that the network controllability framework can be applied to SD models to facilitate the discovery of model levers, i.e., effective leverage points, in the model analysis phase. More specifically, we use network controllability to identify the minimum number of driver nodes N_D (variables) in SD models that is sufficient, if handled appropriately, to exert full model control according to network theory. This is relevant for policy design because the detection of high-leverage points is still an exceedingly difficult task in complex SD models. This study follows the path opened up by Moschoyiannis [100] and Penn et al. [111] who successfully applied network controllability to fuzzy cognitive maps, a modeling field related to SD.

So how can the analytical tools of network controllability be applied to SD models in practice? Essentially, an SD model can be imagined as a web of interrelated causal factors that are assumed to give rise to the complex policy problem under study. Due to its web similarity, the structure of an SD model can be accurately described as a directed weighted network (weighted digraph), making it accessible to algorithmic exploration using concepts from the fields of graph theory and network science [71, 105, 118]. This implies that variables and causal relationships in SD models can be translated into vertices connected by edges. Once an SD model is converted to a network (graph) representation, the application of the network controllability framework is straightforward.

Thus, we conceive the combination of SD and network controllability as a powerful formal analysis method that complements well established tools such as pathway participation metric (PPM), model structure analysis (MSA) or EEA. Figure 5.1 shows how the analytical tools of network controllability fit into the large scheme of formal analysis methods in SD. Importantly, while EEA methods and PPM link model structure to model behavior, MSA and network controllability are limited to characterizing model structure only. Obviously, both MSA and network controllability enable a less nuanced model analysis compared to the other two but they are clearly superior in case of qualitative model analysis since there behavioral information is absent.

The article is structured as follows: We first introduce the main concepts of network controllability and discuss the mathematical procedure to derive the minimum number of driver nodes N_D in an arbitrary complex directed network that is needed to steer the entire network to any state within finite time. As typically multiple driver node configurations of size N_D exist, in a next step, we describe two further node classification schemes. We then illustrate the network controllability approach using the World Dynamics model, i.e., World2 [40] and discuss the potential benefits of integrating network controllability into SD for

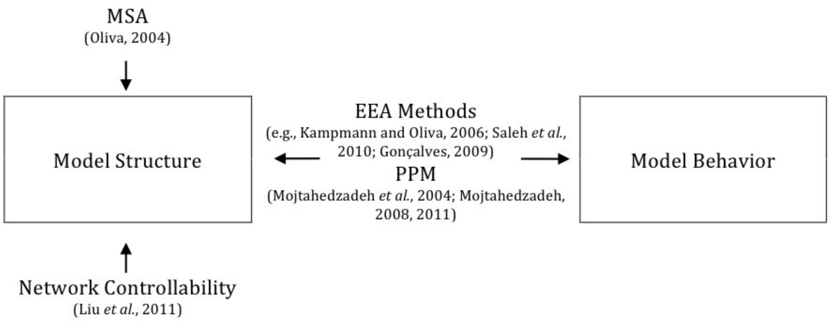


Fig. 5.1 Contribution of network controllability to formal model analysis in SD.

system dynamicists. We conclude by summarizing our results and provide recommendations for future research.

5.2 Network controllability

A system is said to be controllable if we can steer it from any initial state to any desired final state in finite time [70]. Controllability can be easily illustrated with stick balancing, i.e., to balance a stick on a palm. From our experience, we know that this is possible, implying that the system must be controllable [91]. In general controllability is a precondition of control, thus understanding the topology of the underlying network that determines a system's controllability provides numerous insights into the control principles of complex systems [86]. The approach considered in this article is based on the linear time-invariant control system:

$$\dot{x}(t) = \mathbf{A}x(t) + \mathbf{B}u(t). \tag{5.1}$$

where $x(t)$ is a column vector representing the state of the N nodes at time t, $\mathbf{A} := (a_{ij})_{N\times N}$ is the state matrix capturing the weighted wiring diagram of the underlying network, $\mathbf{B} := (b_{im})_{N\times M}$ is the input matrix identifying the nodes that are directly controlled, and u(t) is an input vector. Additionally, a_{ij} is the strength or weight with which node j influences node i where a positive (negative) a_{ij} means the edge $j \leftarrow i$ is excitatory (inhibitory) and $a_{ij} = 0$ if node j has no direct influence on node i; b_{im} represents the strength of an external control signal $u_m(t)$ injected into node i. The linearized system in equation 5.1 is controllable if and only if the $N \times NM$ controllability matrix $\mathbf{C}$ has full rank, i.e.:

$$rank(\mathbf{C}) = N. \tag{5.2}$$

where $\mathbf{C} := [B, AB, A^2B, \cdots, A^{(N-1)}B]$ [70]. If equation 5.2 is satisfied, then we can find an appropriate input vector $u(t)$ to steer the system from any initial state $x(0)$ to an arbitrary final state $x(t)$, implying that the system is controllable [87]. From the definition of the controllability matrix **C** it becomes clear that the network topology, captured by **A**, has a significant impact on controllability. In large networks, however, calculating **C** is computationally demanding and often the system parameters, i.e., the elements in A, are not precisely known. To circumvent the latter problem, [87] use structural controllability [85] where **A** and **B** are considered structured matrices, i.e., their elements are either fixed zeros or independent free parameters. The system in equation 5.1 is structurally controllable if we can fix the nonzero elements in **A** and **B** such that the resulting system satisfies equation 5.2. This has the advantage that we can perform the controllability test described in equation 5.2 even in the absence of complete knowledge of all edge weights a_{ij} in the network [87].

Any network is entirely controllable if we control each node individually. However, in practice this is almost always not feasible and thus we are interested in identifying the smallest subset of nodes, i.e., the minimum number of driver nodes N_D, that when steered by different input signals, can offer full control over the network. In other words, we want to control a network with minimal inputs [86]. Equation 5.2 will not help in finding N_D because it only tells if a network is controllable or not. However, it can be shown that identifying N_D is equivalent to the maximum matching of the network, a purely graph theoretical problem [87, 90]. The maximum matching is the maximal set of edges in a network (graph) that do not share common nodes. A node is considered matched if there is an edge in the maximum matching set that points to it. It has been proven that we can gain full control over a directed network if and only if we directly control each unmatched node and directed paths from the input signals to all matched nodes exist [87].

Thus, to fully control a directed network $G(\mathbf{A})$, the minimum number of driver nodes N_D, is

$$N_D = max\{N - |M*|, 1\}. \tag{5.3}$$

where $|M*|$ is the size of the maximum matching in $G(\mathbf{A})$, i.e., the number of matched nodes. Put differently, the minimum number of driver nodes N_D in a network can be determined from the number of unmatched nodes $N - |M*|$. In the limit case when all nodes are matched ($|M*| = N$) only one input is needed to control the entire network, i.e., $N_D = 1$. A maximum matching of a directed network can be efficiently found using the Hopcroft-Karp algorithm [62]. However, as there might be multiple maximum matchings for a directed network $G(\mathbf{A})$, so can multiple driver node configurations exist, all of size N_D, that can be used for network control. For this reason, to better characterize the role of individual nodes in control, network scientists developed several node classification schemes. Jia et al.

[66] suggest classifying nodes according to their probability of being included in a driver node configuration. A node is: 1) *critical*, if that node must always be controlled to control the system, implying that it is part of all driver node configurations; 2) *intermittent*, if it is a driver node in some driver node configurations but not in all; 3) *redundant* if it is never required for control, implying that it is not part of any driver node configuration.

Alternatively, Liu et al. [88] introduced control centrality to quantify the ability of a single node in controlling an entire network. Centrality measures, i.e., tools to measure the relative importance of nodes, have a long tradition in network research. Depending on the research context, centrality measures such as degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, PageRank, hub and authority centrality, and routing centrality have proven useful. Referring to these classic centrality measures, Schoenenberger and Schenker-Wicki [118] presented a first attempt to apply them to an SD model. In mathematical terms, the control centrality of node i captures the dimension of the controllable subspace or the size of the controllable subsystem when we control node i only. This can be measured with the rank of the controllability matrix **C**, defined as $rank(\mathbf{C})$, indicating the dimension of the controllable subspace of the linearized system in equation 5.1. So when we control node i only, the input matrix B reduces to the vector b^i with a single non-zero entry, and **C** becomes $\mathbf{C}^i$. Similar to before, when the exact value of the edge weights is not entirely known, **A** and **B** are considered structured matrices. In this case, the size of the controllable subspace is measured using the generic rank, ($rank_g$), of $\mathbf{C}^i$ [67, 88]. Consequently, the control centrality of a node i, i.e., $C_c(i)$, is defined as

$$C_c(i) := rank_g(\mathbf{C^i}). \tag{5.4}$$

If $rank_g(\mathbf{C^i} = N)$, then node i alone can control the entire network, i.e., it can steer the network between any points in the N-dimensional state space in finite time. Any value of $rank_g(\mathbf{C^i} = N)$ less than N specifies the dimension of the subspace i can control. Particularly, if $rank_g(\mathbf{C^i} = N)$=1 node i can only control itself [88]. Equation 5.4 can also be normalized as follows:

$$C_c(i) := \frac{rank_g(\mathbf{C^i})}{N}. \tag{5.5}$$

5.3 Application of network controllability to the world dynamics model

For network controllability to be applied, the World Dynamics model [40] needs to be translated into a directed network. First, we have slightly simplified Forrester's model by eliminating the lookup variables (time tabs) from the model. The resulting model contains 64 nodes and 93 edges (see Table C.1 in the Appendix for the list of variables used). Second, we encoded the World Dynamics model in the form of its standard adjacency matrix **A**, i.e., the state matrix in equation 5.1. **A** is an $N \times N$ square matrix that stores information about both the number of nodes and the exact location of all edges between them (Newman, 2013). In this case, **A** has $4096 (64 \times 64)$ entries where $a_{(i,j)} = 1$ if an edge from i to j exists and $a_{(i,j)} = 0$ otherwise.

Then, we determined: 1) if the network (i.e., the structure of the World Dynamics model) is controllable by checking if equation 5.2 is satisfied; 2) the minimum number of driver nodes N_D that offers full control over the network; 3) the node assignment based on the node classification scheme introduced previously [66]; and 4) the control centrality $c_c(i)$ for every node in the World Dynamics model [88]. Finally, in a basic experimental set-up, we show that nodes with high control centrality indeed have a more substantial impact on model behavior than nodes with low control centrality. To perform the analysis we used the tools developed in [87, 88]. Table 5.1 summarizes the procedural steps for the application.

Table 5.1 Procedural steps for the application of network controllability to an SD model (here the World Dynamics model).

Step	Description
1	Preprocess the SD model. Codify the SD model into its standard adjacency matrix **A**. To simplify the coding, in a manual step, variables with no real meaning for the complex policy problem under study, i.e., lookup variables or time constants, can be omitted.
2	Use the standard adjacency matrix **A** derived in step (1) as an input for the controllability analysis. The analysis can be done using the following C++ code packages: 'ControllabilityAnalysis' [87] and 'CalControlCentrality' [88]. The code is available under https://scholar.harvard.edu/yyl/code; creator permission is necessary
3	Perform the analysis in order to answer the following questions: 1) Is the structure of the SD model (network) controllable at all?; 2) How many and which variables are sufficient to exert full control over the structure of the SD model given it is controllable?; 3) How important are individual variables (nodes) in controlling the structure of the SD model?

The analysis yielded that the network under study is indeed controllable. The minimum number of driver nodes N_D equals 34 implying that it is enough to control only 53% of all nodes, i.e., $N_D/N * 100$, to steer the entire network to any point in the N-dimensional state space. Figure 5.2 displays all nodes, i.e., variables, in the World Dynamics model and their classification into critical, intermittent, and redundant nodes is highlighted using different shapes. To avoid a 'crowded' Figure 5.2, we had to abbreviate the node names. An exhaustive variable list with both full names and abbreviations can be found in Appendix C.1. All critical nodes in the World Dynamics model, highlighted as shaded squares, are parameters meaning that they belong to all driver node configurations (of size N_D). This seems intuitive to system dynamicists, because it is the parameters that are their primary target when it comes to policy analysis. From a network controllability perspective, this result is not surprising since all nodes having no incoming links, i.e., the exogenous variables (parameters), must be directly controlled [117].

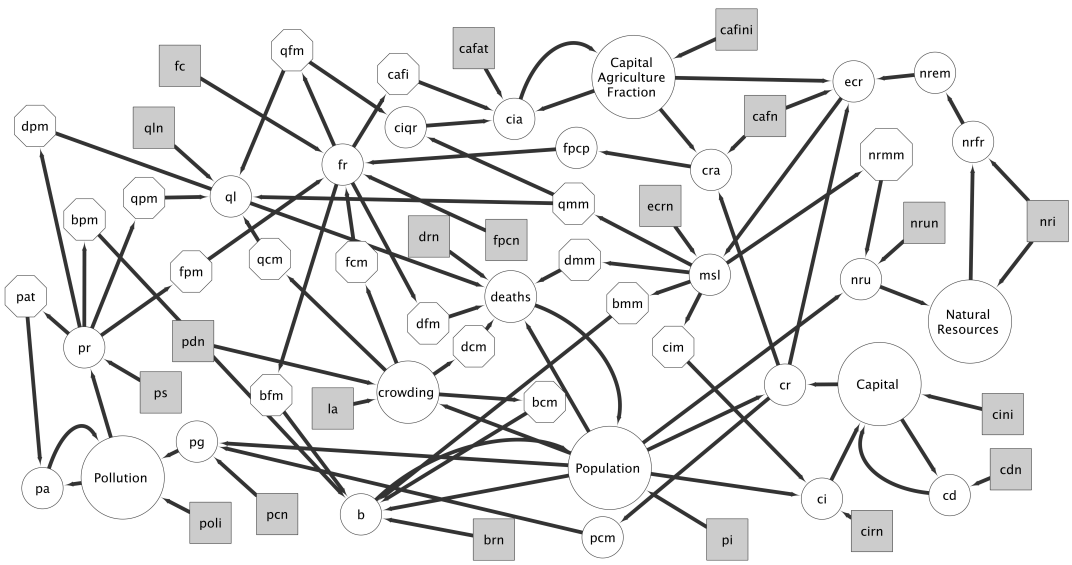


Fig. 5.2 Classification of nodes based on their roles in control in the World Dynamics model.

The intermittent nodes in the World Dynamics model, highlighted as hexagons, are a subset of all the variables (auxiliaries) in the model. Interestingly, these variables are, with two exceptions, i.e., pollution absorption time (pat) and capital agriculture fraction indicated (cafi), multipliers in the World Dynamics model. Therefore, from a purely structural viewpoint, these multipliers seem highly relevant in controlling the World Dynamics model. Finally, the redundant nodes in the World Dynamics model, highlighted as circles, comprise all stock and flow variables, and so they do not have to be directly manipulated to control the World Dynamics model. This is consistent with SD practice where stocks cannot be directly controlled but only through their flows which in turn are steered by parameters.

Now we dive deeper into the node classification by analyzing the control capacity of individual nodes. As discussed earlier, the control centrality corresponds directly to our

Node name	Abbr.	Type	Class [66]	$c_c(i)$ [88]
capital.depreciation.normal	cdn	P	critical	0.44
capital.investment.rate.normal	cirn	P	critical	0.44
land.area	la	P	critical	0.44
natural.resource.utilization.normal	nrun	P	critical	0.44
population.density.normal	pdn	P	critical	0.44
capital.investment.multiplier	cim	V	intermittent	0.44
nat.res.matl.multiplier	nrmm	V	intermittent	0.44

(a) Nodes with the highest influence

Node name	Abbr.	Type	Class [66]	$c_c(i)$ [87]
quality.pollution.multiplier	qpm	V	intermittent	0.03
quality crowding multiplier	qcm	V	intermittent	0.03
quality.of.life.normal	qln	P	critical	0.03
quality.of.life	ql	V	redundant	0.02

(b) Nodes with the lowest influence

Table 5.2 **Nodes with high and low influence on the World Dynamics model**. P: Parameters; V: Variables; $c_c(i)$: Normalized control centrality of node i.

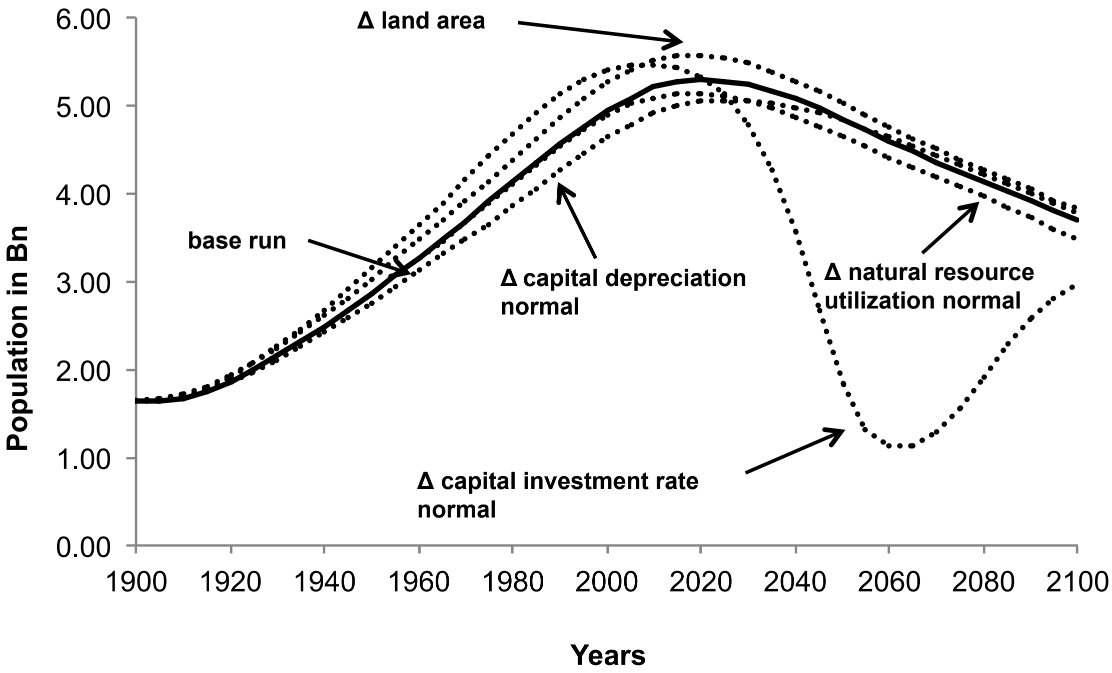
intuition of how powerful a single node is (or groups of nodes are) in controlling the whole network [88]. Table 5.1 shows both the 7 nodes with the highest normalized control centrality $c_c(i)$, all attaining the same score, and the 4 nodes with the lowest $c_c(i)$ in the World Dynamics model. In the model, the range of $c_c(i)$ lies between $[0.02, 0.44]$. We chose to display only the 4 lowest scoring nodes because they have a significantly lower $c_c(i)$ than all the other nodes in the sample (see Table C.1 in the Appendix). The top 7 nodes all achieve a $c_c(i)$ of 0.44 and consist of five parameters and two variables (auxiliaries). In particular, the parameters might serve as effective leverage points in the World Dynamics model.

Now we test if the nodes with a high normalized control centrality, $c_c(i)$, have indeed a more substantial impact on the behavior of the World2 model than the ones with a low $c_c(i)$. To check this, we performed the following test. In the first step we selected the four lowest scoring nodes and compared them against four most influential nodes. Then, in the sense of policy experiments, we separately increased the value of all eight nodes by 10% and ran eight different simulations, i.e., in every simulation only one variable is changed. We assessed the impact these changes have on the five stocks: Population, Capital, Capital Agriculture Fraction, Pollution, and Natural Resources, by visual inspection only. As a reference curve, a base run according to Forrester's [40] original model parameterization is executed. Figure 5.3 shows the impact of a 10% increase of both the four influential nodes, i.e., the ones with the highest $c_c(i)$, and the four ineffective nodes, i.e., the ones with the lowest $c_c(i)$, on the key stock variable in the World2 model: Population. Most notable, individually increasing the four least scoring nodes by 10% has no impact at all on Population and the other stock variables (not shown in Figure 5.3). In contrast, raising the 4 highest scoring nodes by 10% has a significant impact on the trajectory of Population and the other stock variables. Particularly, increasing the parameter capital investment rate normal by 10% not only changes the maximum and final equilibrium point of the trajectory but also its general shape (mode of behavior), i.e., from overshoot and decay to damped oscillation. In conclusion, the test provides a strong indication that high scoring nodes have much more influence on the behavior of the World2 model than low scoring nodes.

5.4 Integration of network controllability into system dynamics

We believe that network controllability is a good complement to the formal analysis techniques in SD (see Figure 5.1). Particularly, we see a significant synergistic potential with MSA [105] which mainly focuses on feedback structures. Based on a purely structural comprehension of SD models, [105] is able to derive the hierarchy of feedback loops in models. In contrast, network controllability concentrates on single nodes and their role in the control of directed networks. In principle, we see two possible options for integrating network controllability into the SD process. The first is the integration of network controllability into model analysis (focus of this article). Alongside other well-established formal analysis techniques, network controllability might serve as a first screening tool of complex SD models for the purpose of identifying leverage points (policy design) within them. The second is the integration of network controllability into model building. Network controllability has the potential to guide the model building process. It is probably most useful when small models are expanded to medium sized ones or when qualitative conceptual maps are transformed into working simulation models. This is because network controllability helps to focus model building on variables that are crucial to the complex policy problem under study. In this context, network controllability might support system dynamicists in defining model regions that are worthwhile to expand or in defining key variables that need to be parameterized for a quantitative simulation model.

Figure 5.4 illustrates the standard SD process and its possible interfaces to network controllability. In this paper, we illustrated the potential of integrating network controllability into model analysis. This can serve as a preliminary screening tool to identify potential



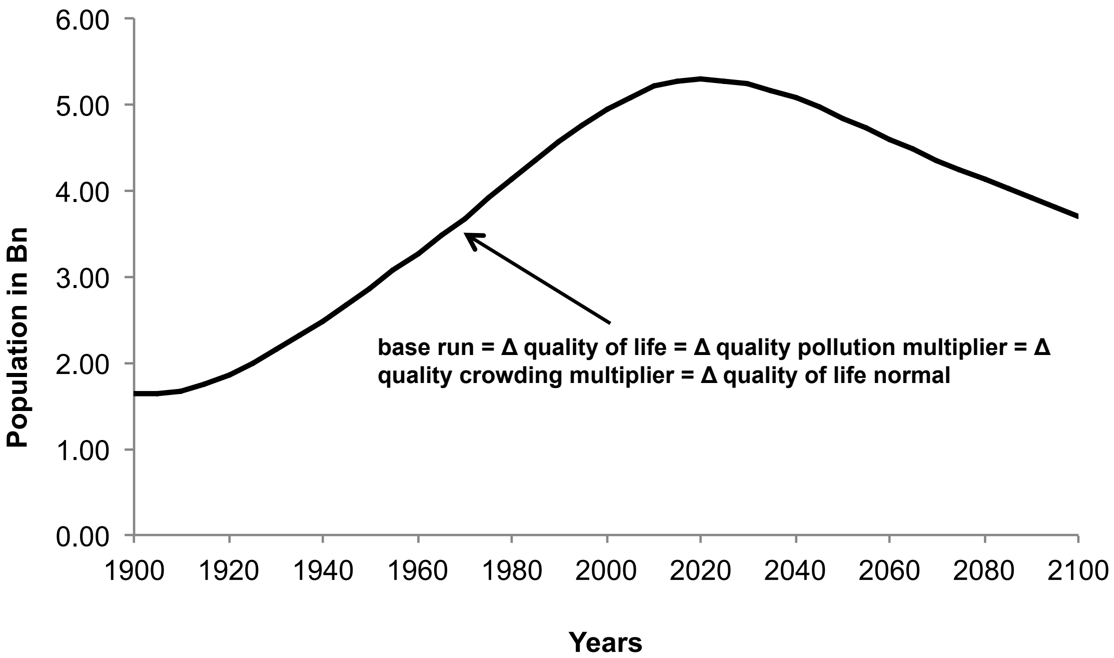


Fig. 5.3 Impact of a 10% increase in the 4 top scoring nodes (upper diagram) and in the 4 least scoring nodes (lower diagram) on Population.

leverage points in SD models as we have demonstrated on the basis of the World Dynamics model. In other words, such an additional structural analysis can assist system dynamicists in designing alternative policies and structures (step 4 in Figure 5.4).

Traditionally, these alternative policies come from intuitive insights generated in the preceding steps of the SD process, from the experience of the modeler, from people operating in the system of interest, or by an exhaustive automatic testing of parameter changes [41]. Consequently, the development of effective alternative policies is difficult, especially in large models, and so a strategy for preliminary determination of candidate nodes for policy design is very helpful.

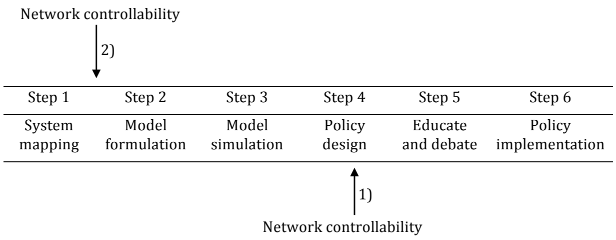


Fig. 5.4 SD process, based on [41], and its interfaces to network controllability.

5.5 Discussion and conclusions

In this article, we argue for an integration of network controllability into the SD process to enhance the current toolset of system dynamicists in formal model analysis. More specifically, we conceive of network controllability as a powerful complementary tool to MSA in exploring the structure of SD models. In contrast to MSA, which deals with feedback complexity, network controllability focuses on single nodes and their role in the control of directed networks. Therefore, network controllability, might be most valuable as a preliminary screening tool of complex SD models to detect potential leverage points within them (see step 4 in Figure 5.4). Every modeler is confronted with two key challenges: how to best represent or model the system, and where to change the system to generate more favorable system outcomes. We believe that network controllability can help modelers address the latter problem by providing such a screening tool.

Merging SD with network controllability is a new approach and has limitations that prescribe future research avenues. First, network controllability applied to SD offers a less nuanced analysis than EEA methods or PPM because it is limited to exploring model structure only. It is clear that system dynamicists are most interested in system behavior and not in structure per se. However, one of the key pillars of SD emphasizes that system behavior arises from underlying system structure [105, 95]. Second, while network controllability provides information about which nodes modelers must tackle for full network control, it does not say how and how much nodes have to be changed. As a consequence, nodes might have to be changed by so much that it is infeasible to implement this change in practice. Third, so far we have only shown the effectiveness of control central nodes in steering model behavior in a basic experimental setting. Thus, future research should be directed towards a systematic investigation of the effect of control central nodes on model behavior by evaluating multiple SD models. Fourth, network controllability builds on the strong assumption that linearizing a nonlinear dynamic system near its equilibrium point is a reasonable procedure. In EEA, so far, this assumption has served well but it is unclear if this holds for the combination of network controllability with SD as well. It is certain, however, when model nonlinearities begin to be dominant determinants of model behavior, the value of linear analysis is limited [37].

Finally, a modeler might often be interested to tune (refine) a model sector only and thus to concentrate the analysis on one specific model region. The methodology we have described so far can only accommodate such cases when the sub-model of interest can be treated as an independent part of the rest of the model. When this is the case, one can reduce the standard adjacency matrix $\mathbf{A}$ to an L dimensional matrix, where L represents the number of variables in the sub-model, and use this as an input to the controllability analysis. However, such cases are rarely encountered in practice and thus an extension of the current methodology to account for controlling sub-networks while using the entire network structure might provide an important avenue for future research.

Chapter 6

Conclusions

In this thesis, we studied social influence processes from three, distinct, angles. In the first study, we proposed a method to identify influential individuals from time-varying social interactions using the wisdom of the crowds principle. To identify influencers, most extant work we described in Chapter 2 advances a two-step approach. In the first step, researchers aim to identify a set of features, typically related to personality, knowledge or position in the social network, which make individuals influential. Once the set of features has been constructed, in the second step the influencers are identified as the individuals with the highest values of the features. While this is a powerful approach, with notable results across a wide range of applications (e.g., [77, 51, 65]), by construction, it suffers from one major limitation: it relies on the assumed relevance of the selected features. The conflicting results reported in recent literature (see Chapter 2) suggest there is no unique set of features describing influencers, which can be explained by influence being context dependent. To integrate this into the influencer identification, the two-step framework would need to be extended by incorporating one more dimension: the context of the social interactions.

Rather than pursuing this path, which could lead to an overly-complicated framework, we propose an approach to identify influencers based on the reaction of the social group to their actions. We develop a framework in which individuals in the social group repeatedly evaluate the contribution of other members according to what they perceive is important and propose a method to aggregate the individual evaluations into a collective judgment. Following this approach, we do not make assumptions on what are the relevant features of the influencers, but we let each individual decide on his own, based on the preferences and beliefs held in that context and at that point in time. The contribution of our study is threefold. First, the results of analyzing more than 50 Mil. posts across three news discussion forums show that influence is indeed context-dependent, with very few individuals being influential across more than one discussion topic. Second, by operationalizing influence in terms of behavioral reactions, we can directly observe the link between the actions of the influencers and the reactions of the individuals in the social group, circumventing the identification problem present in most social influence studies (see [131] for an exception). Third, we provide an identification method which is computationally scalable and thus can be used in real-time applications where immediate actions are necessary.

While our work allows identifying influential individuals in different contexts, it does not explain why they are influential. Understanding what triggers the observed behavioral reactions and which are the consequences of different triggers for the diffusion process will provide an important avenue for further research. Furthermore, our results show that in all three datasets the individuals in the social group agree to a large extent on which are the valuable contributions, which in turn facilitates the influencer identification. However, this might not always be the case. Online discussions can be affected by several processes, including: relevance sorting (posts which receive more votes are more likely to be shown first); order bias (people tend to read only the posts on top of the list) or confirmation bias (people read only information that resonates with their current knowledge). An interesting question for further research is to study if mitigations of such effects through platform design (e.g. display the posts in chronological order vs. most voted on top; include more/less personalization in the ranking algorithms) can have a substantial impact on the emergence of influencers.

In the second study, we investigated how social influence affects repurchase decisions in consumption situations characterized by variety seeking behavior. A close look at the literature reviewed in Chapter 2 reveals that most frequently, social influence is discussed in relationship with a form of *change* in attitudes, opinions or behavior. The Dictionary of Personality and Social Psychology [56] (as cited in [6]) even defines social influence as "any change which a person's relations with other people (individual, group, institution or society) produce on his (sic) intellectual activities, emotions or actions". In consequence, for most existing frameworks, social influence is an agent of change. The primary objective of the second study is to question this one-sided view and show that social influence can have a substantial effect on repeated behavior. To illustrate this, we consider a repurchase decision is the result of a two-step latent process. In the first step, the individual makes a decision based on intrapersonal stimuli, out of which we focus on variety seeking behavior. In the second step, the first step decision is confronted with the social group and thus subjected to social influence. This process can result in two types of social influence processes: (1) visible influence, which determines a change in attitudes, opinions or behavior when there is no inherent intention to change; and (2) hidden influence, which determines preserving the existing attitudes, opinions or behavior against the inherent intention to change. The results from two controlled experiments and an empirical study bring supporting evidence for both processes, showing that hidden influence has a positive effect on the probability to repurchase, while visible influence and variety seeking have a negative effect. We explore these effects by constructing a theoretical Markov model of product choice. We show that when both variety seeking and social influence have a positive valence towards product switch, the effect of social influence on product switch might be overestimated as the decision to switch is already made by the consumer and thus social influence has merely a reinforcement effect. On the other hand, when variety seeking has a positive valence towards product switch and social influence has a positive valence towards repurchase, a repurchase reflects an actual influence effect, as in the absence of social influence the individual would have switched to the new product. Our work contributes to existing literature by providing a new conceptualization of social influence and illustrating its positive effect on repurchase behavior in consumption situations characterized by variety seeking behavior. This is to our knowledge the first study documenting such an effect and shows that individuals who do not promote opinion or behavioral change can be as important for diffusion processes as are those who promote it. While our study brings supporting evidence to the main effects, it does not explore interaction effects nor the role of potential moderators, like product price. Does variety seeking have the same effect for low vs. high values of the social signal? Furthermore, are people more inclined to give up their variety seeking to social influence for more versus less expensive products? Understanding such interaction effects might provide a significant contribution of future research. A different path that could be followed in subsequent work is to study the relationship between variety seeking and susceptibility to social influence. High variety seeking individuals switch products more often, but this does not imply they are more likely to do this following a social influence attempt. Thus an important question that can be raised is if high variety seeking individuals are more (or less) susceptible to social influence compared to low variety seeking individuals. If high variety seeking individuals are less susceptible to social influence than low variety seeking individuals, acquiring them as customers is both demanding and not effective, as they will soon change to a different product due to variety seeking. If they are more susceptible, the ease of acquiring them might be a good trade-off for their relatively low expected customer lifetime value.

In the third study, we showed that influencer identification tools have a broad application and can be used to answer important questions outside the context they were designed for. To this end, we illustrate how tools for influencer identification can facilitate the design of complex policies. More specifically, we show how effective leverage points (i.e., model levers) in system dynamics models representing complex policy problems can be identified using network controllability. We propose a framework where, in the first step, we map a complex policy problem to a directed network, with nodes representing variables and links the causal relationships between them. In the second step, we identify the model levers as the most influential nodes in steering the entire system to an arbitrary final state. We apply our approach to a classical system dynamics model: the World dynamics model [40] and show that changing the most influential variables (i.e., the identified model levers) leads to a significantly larger change in the variables of interest compared to changing the least influential ones. We contribute to existing literature by providing an approach to facilitate policy design. The method we developed can be used as a preliminary screening tool of complex system dynamics models to detect effective leverage points. While the method provides information about which variables modelers should focus on for full model control, one limitation is that it does not provide any information about (1) how the system could be steered to one specific state and (2) which would be the cost of doing so. It can happen that the identified variables would have to be changed so much that the change is not feasible to implement in practice. Developing the method to account for the cost required by potential changes might provide an important avenue for further research.

References

[1] In 2017 Influencer Marketing Is About To Go Through The Roof. https://www.inc.com/nicolas-cole/in-2017-influencer-marketing-is-about-to-go-through-the-roof.html. Accessed: 2018-01-16.

[2] Influencer Marketing Study. https://blog.tomoson.com/influencer-marketing-study/. Accessed: 2018-01-16.

[3] Influencer Marketing Update: Non-Celebrity Influencers 10 Times More Likely to Drive In-Store Purchases. https://collectivebias.com/blog/2016/03/influencer-marketing-update-non-celebrity-influencers-10-times-likely-drive-store-purchases/. Accessed: 2018-01-16.

[4] New research: The value of influencers on Twitter. https://blog.twitter.com/marketing/en_us/a/2016/new-research-the-value-of-influencers-on-twitter.html. Accessed: 2018-01-16.

[5] Why YouTube Stars Are More Influential Than Traditional Celebrities. https://www.thinkwithgoogle.com/consumer-insights/youtube-stars-influence/. Accessed: 2018-01-16.

[6] Abrams, D. and Hogg, M. A. (1990). Social Identification, Self-Categorization and Social Influence. *European Review of Social Psychology*, 1(1):195–228.

[7] Algesheimer, R., Borle, S., Dholakia, U. M., and Singh, S. S. (2010). The impact of customer community participation on customer behaviors: An empirical investigation. *Marketing science*, 29(4):756–769.

[8] Algesheimer, R., Dholakia, U. M., and Herrmann, A. (2005). The social influence of brand community: Evidence from european car clubs. *Journal of marketing*, 69(3):19–34.

[9] Anderson, E. G. (2014). Presidential address. *Proceedings of the 2014 International System Dynamics Conference*.

[10] Aral, S. and Walker, D. (2011). Viral Product Design : A Randomized Trial of Peer Influence in Networks. *Management Science*, 57(9):1623–1639.

[11] Aral, S. and Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, 341:337–341.

[12] Ariely, D. and Levav, J. (2000). Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of consumer Research*, 27(3):279–290.

[13] Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).

[14] Barabási, A.-L. (2016). *Network science*. Cambridge university press.

[15] Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286:509–512.

[16] Barlas, Y. (2016). Editorial. *System Dynamics Review*, 32(1):3—-5.

[17] Bass, B. M. (1949). An Analysis of the Leaderless Group Discussion. *Journal of Applied Psychology*, 33.6:527–533.

[18] Baumgartner, H. and Steenkamp, J.-B. E. (1996). Exploratory consumer buying behavior: Conceptualization and measurement. *International journal of Research in marketing*, 13(2):121–137.

[19] Becker, J., Brackbill, D., and Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences*, 114(26):E5070–E5076.

[20] Berger, J. and Schwartz, E. M. (2011). What Drives Immediate and Ongoing Word of Mouth? *Journal of Marketing Research*, 48(5):869–880.

[21] Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182.

[22] Brandes, U., Robins, G., McCranie, A., and Wasserman, S. (2013). What is network science? network science, 1 (1), 1–15.

[23] Budescu, D. V. and Chen, E. (2014). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science*, 61(August 2016):140523081120004.

[24] Burnkrant, R. E. and Cousineau, A. (1975). Informational and Normative Social Influence in Buyer Behavior. *Source Journal of Consumer Research*, 2(3):206–215.

[25] Chae, I., Stephen, A. T., and Bart, Y. (2017). Spillover E ects in Seeded Word-of-Mouth Marketing Campaigns. 36(1):89–104.

[26] Chen, X., van der Lans, R., and Phan, T. Q. (2017). Uncovering the Importance of Relationship Characteristics in Social Networks: Implications for Seeding Strategies. *Journal of Marketing Research*, 54(2):187–201.

[27] Chevalier, J. A. and Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3):345–354.

[28] Cialdini, R. B. and Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, 55(1):591–621.

[29] Coleman, J., Katz, E., and Menzel, H. (1977). The diffusion of an innovation among physicians. *Social networks: A developing paradigm*, pages 107–124.

[30] Coleman, J. S., Katz, E., and Menzel, H. (1966). Medical innovation: A diffusion study.

[31] Colman, A. M. (2015). *A dictionary of psychology*. Oxford University Press, USA.

[32] Corey, L. G. (1971). People who claim to be opinion leaders: identifying their characteristics by self-report. *The Journal of Marketing*, pages 48–53.

[33] Deci, E. L. and Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. In *Advances in experimental social psychology*, volume 13, pages 39–80. Elsevier.

[34] Deutsch, M. and Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3):629.

[35] Dholakia, U. M., Bagozzi, R. P., and Pearo, L. K. (2004). A social influence model of consumer participation in network- and small-group-based virtual communities. *International Journal of Research in Marketing*, 21(3):241–263.

[36] Du, R. Y. and Kamakura, W. A. (2011). Measuring contagion in the diffusion of consumer packaged goods. *Journal of Marketing Research*, 48(1):28–47.

[37] Eberlein, R. L. (1989). Simplification and understanding of models. *System Dynamics Review*, 5(1):51–68.

[38] Elwert, F. (2015). Comment Public health : real-world network targeting of interventions. *The Lancet*, 6736(15):1–2.

[39] Forrester, J. W. (1969). Urban dynamics.

[40] Forrester, J. W. (1971). *World dynamics*. Wright-Allen Press.

[41] Forrester, J. W. (1994). System dynamics, systems thinking, and soft or. *System dynamics review*, 10(2-3):245–256.

[42] Fossen, B. L. and Schweidel, D. A. (2017). Television Advertising and Online Word-of-Mouth: An Empirical Investigation of Social TV Activity. *Marketing Science*, 36(1):105–123.

[43] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(1968):215–239.

[44] Galton, F. (1907). Vox populi. *Nature*, 75(7):450–451.

[45] Gao, J., Liu, Y.-Y., D'souza, R. M., and Barabási, A.-L. (2014). Target control of complex networks. *Nature communications*, 5:5415.

[46] Ghaffarzadegan, N., Lyneis, J., and Richardson, G. P. (2011). How small system dynamics models can help the public policy process. *System Dynamics Review*, 27(1):22–44.

[47] Gitlin, T. (1978). Media sociology. *Theory and society*, 6(2):205–253.

[48] Givon, M. (1984). Variety seeking through brand switching. *Marketing Science*, 3(1):1–22.

[49] Godes, D. (2011). Commentary—invited comment on “opinion leadership and social contagion in new product diffusion”. *Marketing Science*, 30(2):224–229.

[50] Godes, D. and Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science*, 28(4):721–739.

[51] Goldenberg, J., Han, S., Lehmann, D. R., and Hong, J. W. (2009). The Role of Hubs in the Adoption Process. *Journal of Marketing*, 73:1–13.

[52] Goldenberg, J., Lehmann, D. R., Shidlovski, D., and Barak, M. M. (2006). The Role of Expert versus Social Opinion Leaders in New Product Adoption. *MSI Reports*, (06-124).

[53] Granovetter, M. S. (1977). The strength of weak ties. In *Social networks*, pages 347–367. Elsevier.

[54] Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing research*, pages 342–355.

[55] Hamilton, R. W., Schlosser, A., and Chen, Y.-J. (2017). Who’s Driving This Conversation? Systematic Biases in the Content of Online Consumer Discussions. *Journal of Marketing Research*, 54(4):540–555.

[56] Harré, R. and Lamb, R. (1986). The dictionary of personality and social psychology.

[57] Heckman, J. J. and Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, 9(2):85–110.

[58] Hinz, O., Skiera, B., Barrot, C., and Becker, J. U. (2011). Social Contagion – An Empirical Comparison of Seeding Strategies for Viral Marketing. *Journal of Marketing*, 75(November):55–71.

[59] Holme, P. (2015). Modern temporal network theory: A colloquium. *European Physical Journal B*, 88(9):1—-30.

[60] Homer, J., Hirsch, G., Minniti, M., and Pierson, M. (2004). Models for collaboration: How system dynamics helped a community organize cost-effective care for chronic illness. *System Dynamics Review*, 20(3):199–222.

[61] Honggang, X., Mashayenkhi, A., and Saeed, K. (1998). Effectiveness of infrastructure service delivery through earmarking: the case of highway construction in china. *System Dynamics Review*, 14(2-3):221–255.

[62] Hopcroft, J. E. and Karp, R. M. (1973). An nˆ5/2 algorithm for maximum matchings in bipartite graphs. *SIAM Journal on computing*, 2(4):225–231.

[63] Hu, Y. and Van den Bulte, C. (2014). Nonmonotonic Status Effects in New Product Adoption. *Marketing Science*, 33(4):509–533.

[64] Iyengar, R., Van den Bulte, C., and Lee, J. Y. (2015). Social contagion in new product trial and repeat. *Marketing Science*, 34(3):408–429.

[65] Iyengar, R., Van den Bulte, C., and Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(February 2016):195–212.

[66] Jia, T., Liu, Y.-Y., Csóka, E., Pósfai, M., Slotine, J.-J., and Barabási, A.-L. (2013). Emergence of bimodality in controlling complex networks. *Nature communications*, 4.

[67] Johnston, R., Barton, G., and Brisk, M. (2007). Determination of the generic rank of structural matrices. *International Journal of Control*, 40(2):257–264.

[68] Kahn, B. E. and Raju, J. S. (1991). Effects of price promotions on variety-seeking and reinforcement behavior. *Marketing Science*, 10(4):316–337.

[69] Kailath, T. (1980). *Linear systems*, volume 156. Prentice-Hall Englewood Cliffs, NJ.

[70] Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 1(2):152–192.

[71] Kampmann, C. E. (2012). Feedback loop gains and system behavior (1996). *System Dynamics Review*, 28(4):370–395.

[72] Katona, Z., Zubcsek, P., and Sarvary, M. (2011). Network Effects and Personal Influences: Diffusion of an Online Social Network. *Journal of Marketing Research*, XLVIII(June):425–443.

[73] Katz, E. and Lazarsfeld, P. F. (1955). *Personal Influence, The part played by people in the flow of mass communications*. The Free Press.

[74] Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of conflict resolution*, 2(1):51—-60.

[75] Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*, page 137.

[76] Kerr, N. L. (1998). Harking: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3):196–217.

[77] Kim, D. a., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, a. J., Fowler, J. H., and Christakis, N. a. (2015). Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. *The Lancet*, 386(15):145–153.

[78] King, C. W. and Summers, J. O. (1970). Overlap of Opinion Leadership Across Consumer Product Categories. *Journal of Marketing*, VII(1):43–50.

[79] Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. a. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888—-893.

[80] Klemm, K., Serrano, M. Á., Eguíluz, V. M., and San Miguel, M. (2012). A measure of individual role in collective dynamics. *Scientific reports*, 2.

[81] Lambiotte, R. (2016). Rich gets simpler. *Proceedings of the National Academy of Sciences*, 113(36):9961–9962.

[82] Lambrecht, A., Tucker, C., and Wiertz, C. (2018). Advertising to Early Trend Propagators: Evidence from Twitter. *Marketing Science*, 37(14):0–23.

[83] Lawyer, G. (2015). Understanding the influence of all nodes in a network. *Scientific reports*, 5.

[84] Leonard-Barton, D. (1985). Experts as negative opinion leaders in the diffusion of a technological innovation. *Journal of Consumer Research*, 11(4):914–926.

[85] Lin, C.-T. (1974). Structural controllability. *IEEE Transactions on Automatic Control*, 19(3):201–208.

[86] Liu, Y.-Y. and Barabási, A.-L. (2016). Control principles of complex systems. *Reviews of Modern Physics*, 88(3):035006.

[87] Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, 473(7346):167–73.

[88] Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2012). Control centrality and hierarchical structure in complex networks. *Plos one*, 7(9):e44459.

[89] Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22):9020–9025.

[90] Lovász, L. and Plummer, M. D. (2009). *Matching theory*, volume 367. American Mathematical Soc.

[91] Luenberger, D. G. D. G. (1979). *Introduction to dynamic systems; theory, models, and applications*. NY:Wiley.

[92] Luo, X. (2009). Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices. *Marketing Science*, 28(1):148–165.

[93] Luo, X., Andrews, M., Song, Y., and Aspara, J. (2014). Group-Buying Deal Popularity. *Journal of Marketing*, 78(2):20–33.

[94] McAlister, L. and Pessemier, E. (1982). Variety seeking behavior: An interdisciplinary review. *Journal of Consumer research*, 9(3):311–322.

[95] Meadows, D. H. (1989). System dynamics meets the press. *System Dynamics Review*, 5(1):69–80.

[96] Merton, R. K. (1968). *Social theory and social structure*. Simon and Schuster.

[97] Meyners, J., Barrot, C., Becker, J. U., and Goldenberg, J. (2017). The Role of Mere Closeness: How Geographic Proximity Affects Social Influence. *Journal of Marketing*, 81(September):jm.16.0057.

[98] Morone, F. and Makse, H. a. (2015). Influence maximization in complex networks through optimal percolation. *Nature*, 524(08):65—-68.

[99] Morvinski, C., Amir, O., and Muller, E. (2017). "Ten Million Readers Can't Be Wrong!," or Can They? On the Role of Information About Adoption Stock in New Product Trial. *Marketing Science*, 36(2):290–300.

[100] Moschoyiannis, S., Elia, N., Penn, A. S., Lloyd, D. J., and Knight, C. (2016). A web-based tool for identifying strategic intervention points in complex systems. *arXiv preprint arXiv:1608.00655*.

[101] Nair, H. S., Manchanda, P., and Bhatia, T. (2010). Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders. *Journal of Marketing Research*, 47(5):883–895.

[102] Nejad, M. G., Sherrell, D. L., and Babakus, E. (2014). Influentials and influence mechanisms in new product diffusion: an integrative review. *Journal of Marketing Theory and Practice*, 22(2):185–208.

[103] Newman, M. (2010). *Networks: an introduction*. Oxford university press.

[104] Nitzan, I. and Libai, B. (2011). Social effects on customer retention. *Journal of Marketing*, 75(6):24–38.

[105] Oliva, R. (2004). Model structure analysis through graph theory: partition heuristics and feedback structure decomposition. *System Dynamics Review*, 20(4):313–336.

[106] Oliva, R. (2015). Linking structure to behavior using eigenvalue elasticity analysis. *Analytical Methods for Dynamic Modelers, Rahmandad H, Oliva R, Osgood ND (eds). MIT Press: Cambridge, MA*, pages 207–239.

[107] Oliva, R. (2016). Structural dominance analysis of large and stochastic models. *System dynamics review*, 32(1):26–51.

[108] Packard, G. and Berger, J. (2017). How Language Shapes Word of Mouth's Impact. *Journal of Marketing Research*, 54(4):572–588.

[109] Pal, A. and Konstan, J. A. (2010). Expert identification in community question answering: exploring question selection bias. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1505–1508. ACM.

[110] Peng, J., Agarwal, A., Hosanagar, K., and Iyengar, R. (2018). Network Overlap and Content Sharing on Social Media Platforms. *Journal of Marketing Research*, LV(August):jmr.14.0643.

[111] Penn, A. S., Knight, C. J., Chalkias, G., Velenturf, A. P., and Lloyd, D. J. (2017). Extending participatory fuzzy cognitive mapping with a control nodes methodology: A case study of the development of a bio-based economy in the humber region, uk. In *Environmental Modeling with Stakeholders*, pages 171–188. Springer.

[112] Phan, T. and Godes, D. (2018). The Evolution of Influence Through Endogenous Link Formation. *Marketing Science*, 37(2):1–20.

[113] Raju, P. S. (1980). Optimum stimulation level: Its relationship to personality, demographics, and exploratory behavior. *Journal of consumer research*, 7(3):272–282.

[114] Ratner, R. K. and Kahn, B. E. (2002). The impact of private versus public consumption on variety-seeking behavior. *Journal of Consumer Research*, 29(2):246–257.

[115] Risselada, H., Verhoef, P. C., and Bijmolt, T. H. (2014). Dynamic effects of social influence and direct marketing on the adoption of high-technology products. *Journal of Marketing*, 78(2).

[116] Rogers, E. M. (2003). *The Diffusion of Innovations*. The Free Press.

[117] Ruths, J. and Ruths, D. (2014). Control profiles of complex networks. *Science (New York, N.Y.)*, 343:1373–6.

[118] Schoenenberger, L. K. and Schenker-Wicki, A. (2014). Can system dynamics learn from social network analysis? *Proceedings of the 2014 International System Dynamics Conference*.

[119] Schulze, C., Schöler, L., and Skiera, B. (2014). Not All Fun and Games: Viral Marketing for Utilitarian Products. *Journal of Marketing*, 78(1):1–19.

[120] Seiler, S., Yao, S., and Wang, W. (2017). Does Online Word of Mouth Increase Demand? (And How?) Evidence from a Natural Experiment. *Marketing Science*, 36(6):838–861.

[121] Sekara, V., Stopczynski, A., and Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, 113(36):9977–9982.

[122] Simon, H. (1955). On a Class of Skew Distribution Functions. *Biometrika*, 42:425–440.

[123] Slotine, J.-J. E., Li, W., et al. (1991). *Applied nonlinear control*, volume 199. Prentice hall Englewood Cliffs, NJ.

[124] Sridhar, S. and Srinivasan, R. (2012). Social Influence Effects in Online Product Ratings. *Journal of Marketing*, 76(5):70–88.

[125] Stephen, A. T., Zubcsek, P. P., and Goldenberg, J. (2016). Lower Connectivity Is Better: The Effects of Network Structure on Redundancy of Ideas and Customer Innovativeness in Interdependent Ideation Tasks. *Journal of Marketing Research*, 53(2):263–279.

[126] Sterman, J. D. J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. Number HD30. 2 S7835 2000.

[127] Surowiecki, J. (2004). *he Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Doubleday Books, New York.

[128] Tanase, R., Dholakia, U., and Algesheimer, R. (2017a). The effect of opinion leadership on the varied behavior.

[129] Tanase, R., Dholakia, U., and Algesheimer, R. (2017b). The effect of opinion leadership on the varied behavior.

[130] Travençolo, B. a. N. and Da, L. (2008). Accessibility in complex networks. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 373(1):89–95.

[131] Trusov, M., Bodapati, A. V., and Bucklin, R. E. (2010). Determining Influential Users in Internet Social Networks. *Journal of Marketing Research*, 47(August):643–658.

[132] Valente, T. W. (2012). Network Interventions. *Science*, 337(6090):49—-53.

[133] Valente, T. W., Hoffman, B. R., Ritt-Olson, A., Lichtman, K., and Johnson, C. A. (2003). Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. *American journal of public health*, 93(11):1837–1843.

[134] Valente, T. W. and Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health education & behavior : the official publication of the Society for Public Health Education*, 34(X):881–896.

[135] Van den Bulte, C. and Joshi, Y. V. (2007). New Product Diffusion with Influentials and Imitators. *Marketing Science*, 26(February 2016):400–421.

[136] Van Trijp, H. C., Hoyer, W. D., and Inman, J. J. (1996). Why switch? product category: level explanations for true variety-seeking behavior. *Journal of Marketing Research*, pages 281–292.

[137] Van Trijp, H. C. and Steenkamp, J.-B. E. (1992). Consumers' variety seeking tendency with respect to foods: measurement and managerial implications. *European Review of Agricultural Economics*, 19(2):181–195.

[138] Watts, D. J. and Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458.

[139] Weimann, G. (1994). *The Influentials: People Who Influence People*. State University of New York Press.

[140] Yule, G. (1925). A Mathematical Theory of Evolution based on the Conclusions of Dr. J.C. Willis, F.R.S. *Journal of the Royal Statistical Society*, 88:433–436.

[141] Zaleskiewicz, T. (2001). Beyond risk seeking and risk aversion: Personality and the dual nature of economic risk taking. *European journal of Personality*, 15(S1).

[142] Zhang, C., Phang, C. W., Wu, Q., and Luo, X. (2017a). Nonlinear Effects of Social Connections and Interactions on Individual Goal Attainment and Spending: Evidences from Online Gaming Markets. *Journal of Marketing*, 81(6):132–155.

[143] Zhang, Y., Moe, W. W., and Schweidel, D. A. (2017b). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1):100–119.

Appendix A

Identification of influencers through the wisdom of crowds

A.1 $\mathscr{I}(i) \in [0,1], \forall i$

Proposition 1 *Let* $\mathbf{X} = \{x_1, \ldots, x_n\}$ *be a vector with* $x_k \in [0,1], \forall k = 1, \ldots, n$. *Then* $\mathbb{E}[\mathbf{X}] \geq Var(\mathbf{X}), \forall \mathbf{X}$. *Then* $\langle \mathbf{X} \rangle \geq \langle \mathbf{X}^2 \rangle - \langle \mathbf{X} \rangle^2, \forall \mathbf{X}$

Proof 1

$$Var(\mathbf{X}) = \mathbb{E}[\mathbf{X}^2] - \mathbb{E}[\mathbf{X}]^2 \tag{A.1}$$

$$\langle \mathbf{X} \rangle - \langle \mathbf{X}^2 \rangle - \langle \mathbf{X} \rangle^2 \geq 0 \tag{A.2}$$

$$\langle \mathbf{X} \rangle + \langle \mathbf{X} \rangle^2 \geq \langle \mathbf{X}^2 \rangle \tag{A.3}$$

The last inequality always holds as $x_k \in [0,1]$ *and thus* $x_k \geq x_k^2, \forall k = 1, \ldots, n$.

A.2 Small sample bias

The IP is at its root a statistical aggregation and, as any statistical measure, is susceptible to bias arising from small samples. This bias can be induced in two ways: (1) if within an event there are few participants; (2) if an individual takes part in a low number of events. In events with few participants (thus implicitly few judges), the IP scores might be biased as they violate one of the critical assumptions behind the wisdom of crowds: a large number of independent evaluations. To address this we penalise event ranks obtained in small events by introducing the constant term c in the event rank normalisation. By changing c one can emphasise or diminish the role of the event size in computing the IP. Fig. S10 in the SI illustrates the impact c has on computing the event ranks. For large c, high IP values can only be obtained in the limit of large events, while for small c the effect of the event size on the event rank is negligible. This has practical implications for studying dynamical processes like information propagation where the size of the susceptible population plays an important role. The second source of small sample bias is the small number of events attended by an individual. In this case the IP might not be informative for the latent potential to influence as by aggregating few data points the results are subjected to randomness. We can address this by setting a threshold on the minimum number of events attended by each individual and remove from the analysis those who attended less. The threshold can be seen as a measure of confidence in the results. The higher the threshold, the higher is the minimum number of events attended by each individual and thus the lower the likelihood that high vote scores are obtained in most events by chance.

A.3 Preferential attachement null model

To test if data can be explained by the rich-get-richer effect we create a null model in which the probability of a post to receive a vote depends on how many votes the post had so far received. We consider the observed number of votes in a thread was distributed across the posts in the thread at $\tau = 10$ points in time. The time points are select uniformly at random from the time-span of the thread. The number of votes being distributed is the same at all time points (the total number of votes in the thread divided by τ). We compute the probability of a post j written until time s to receive a vote distributed at s as:

$$p(j,s) = \alpha \frac{1}{\sum\limits_{k=1}^{s} q(k)} + (1-\alpha)\frac{v(j,s)}{\sum\limits_{k=1}^{s} v(k)}$$

where $\sum_{k=1}^{s} q(k)$ is the number of posts written until s, $v(j,s)$ the number of votes of post j at time s and $\sum_{k=1}^{s} v(k)$ the total number of votes observed until time s. Thus, at time s, with probability α a vote is given to a post sampled uniformly from all posts written until s and with probability $1-\alpha$ the vote is given to one of the posts written until s sampled with probability determined by preferential attachement. The IP is computed using the sum of randomized vote scores as input. The procedure is repeated 100 times and the IP under the null model is computed as the mean IP over the repetitions.

A.4 Tables and figures

Topic	Abbreviation	NrUsers	NrPosts	NrThreads
us	C1	315650	4493523	4469
world	C2	287436	4649352	8318
opinion	C3	248050	3804811	3221
politics	C4	204574	5567280	2173
justice	C5	164644	2150579	1727
showbiz	C6	116754	707582	2402
tech	C7	97209	528102	1618
health	C8	89191	576744	1456
travel	C9	63477	268125	1411
living	C10	56784	270640	800
sport	-	45632	191879	2727
business	-	24497	86888	1347
studentnews	-	4908	14053	164

Table A.1 CNN dataset. Overview of the topic categories.

Topic	Abbreviation	NrUsers	NrPosts	NrThreads
politics	C1	102490	1564614	9737
business	C2	68169	668984	7287
entertainment	C3	55344	378312	6323
national	C4	50824	580443	4185
international	C5	46334	464150	4978
technology	C6	41968	158793	5066
health	C7	40929	226190	3689
magazine	C8	25675	106026	887
education	C9	15636	120747	954
features	C10	10510	67930	80
sexes	-	9144	43776	466
personal	-	6919	168559	1054
science	-	4366	22310	315
video	-	4048	10712	645
culture	-	3434	91276	543
china	-	2030	7412	369
events	-	726	1943	61

Table A.2 The Atlantic dataset. Overview of the topic categories.

Topic	Abbreviation	NrUsers	NrPosts	NrThreads
news	C1	221315	8713981	51359
business	C2	128747	4012063	68228
sport	C3	114861	2724460	75902
culture	C4	100018	883974	35456
lifestyle	C5	90663	1029662	26773
tech	C6	82976	536461	18165
opinion	C7	51262	2018617	14528
health	C8	48921	508732	10084
education	C9	45714	669610	7890
environment	C10	42462	556758	5591
travel	-	35333	163921	11067
money	-	10964	39099	2639
history	-	9571	55624	458
family	-	6064	23513	749
video	-	793	1495	208

Table A.3 The Telegraph dataset. Overview of the topic categories.

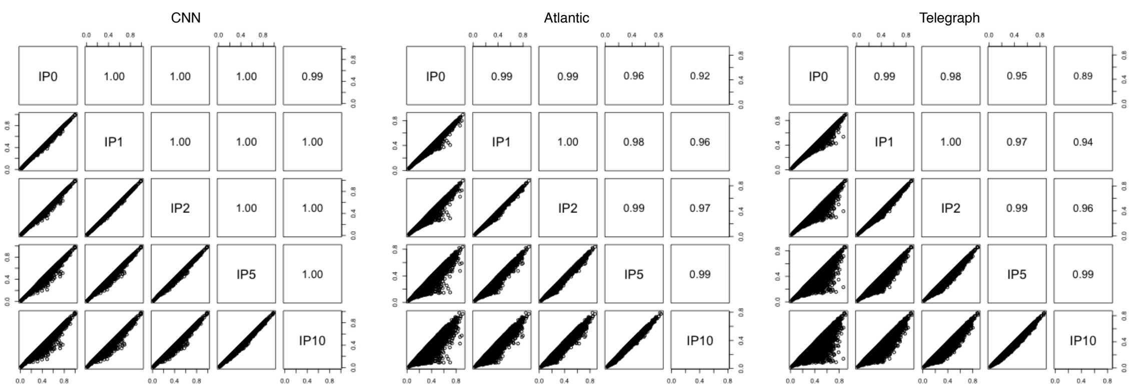


Fig. A.1 **The influence potential is robust to the choice of c**. The diagonal panels show the choice of $c \in \{0,1,2,5,10\}$ used to computed the IP. The lower triangle panels show the pairwise scatterplots between the IP scores computed with different c. The upper triangle panels show the Pearson correlation coefficient. Data is pooled from all topic categories. An individual can be described by multiple data points, each representing his IP in a category where he participated in at least 10 events. For all three datasets there is a very high pairwise correlation (Pearson $r \geq 0.89$) between the IP values for different choices of c.

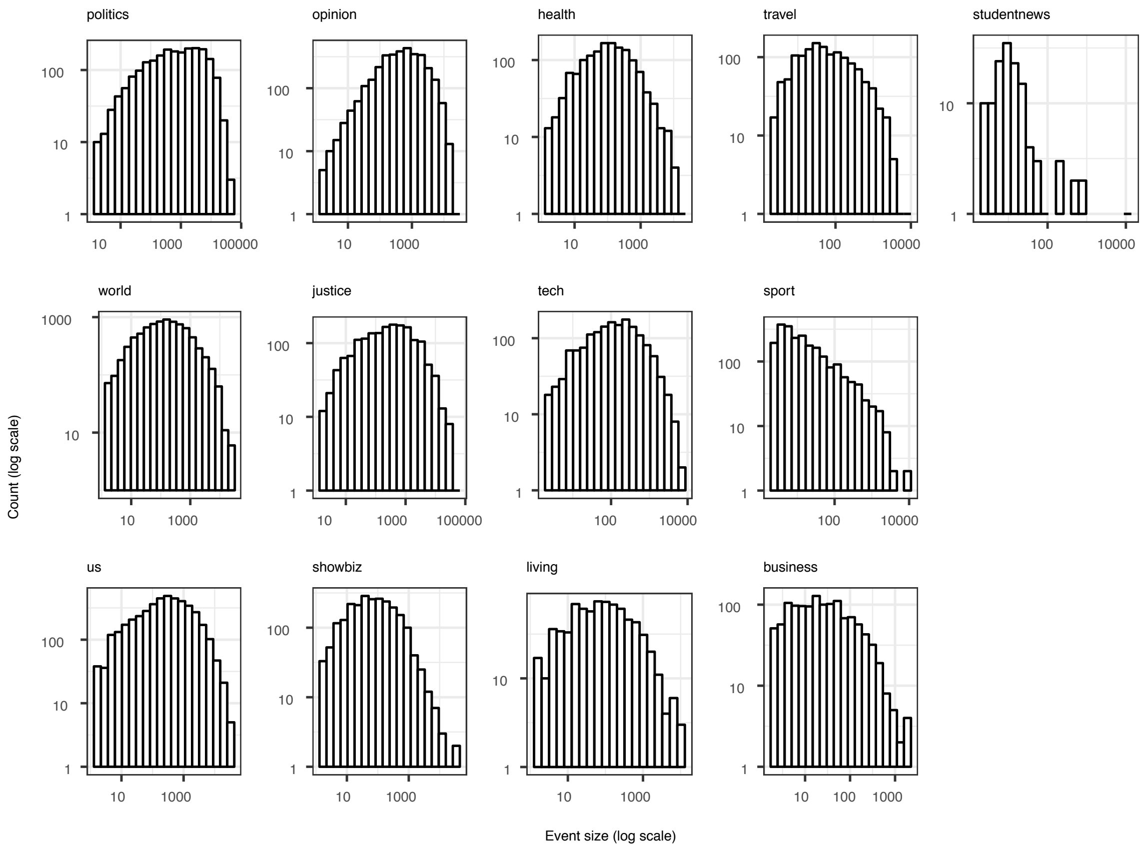


Fig. A.2 Distribution of the event size within the topic categories: CNN dataset.

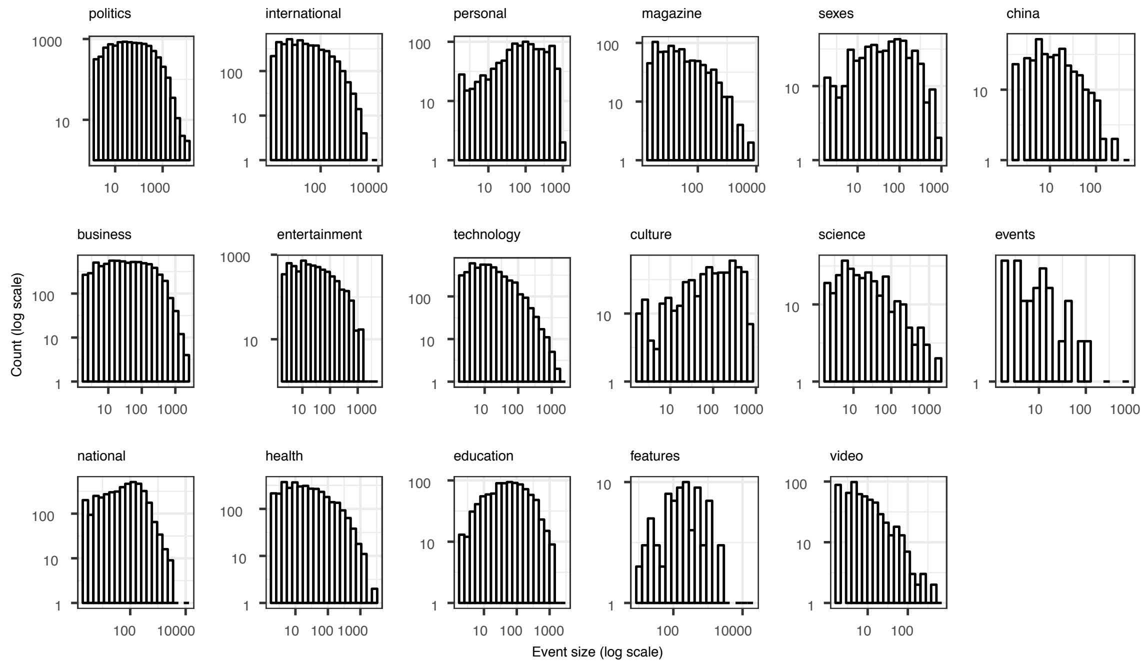


Fig. A.3 Distribution of the event size within the topic categories: Atlantic dataset.

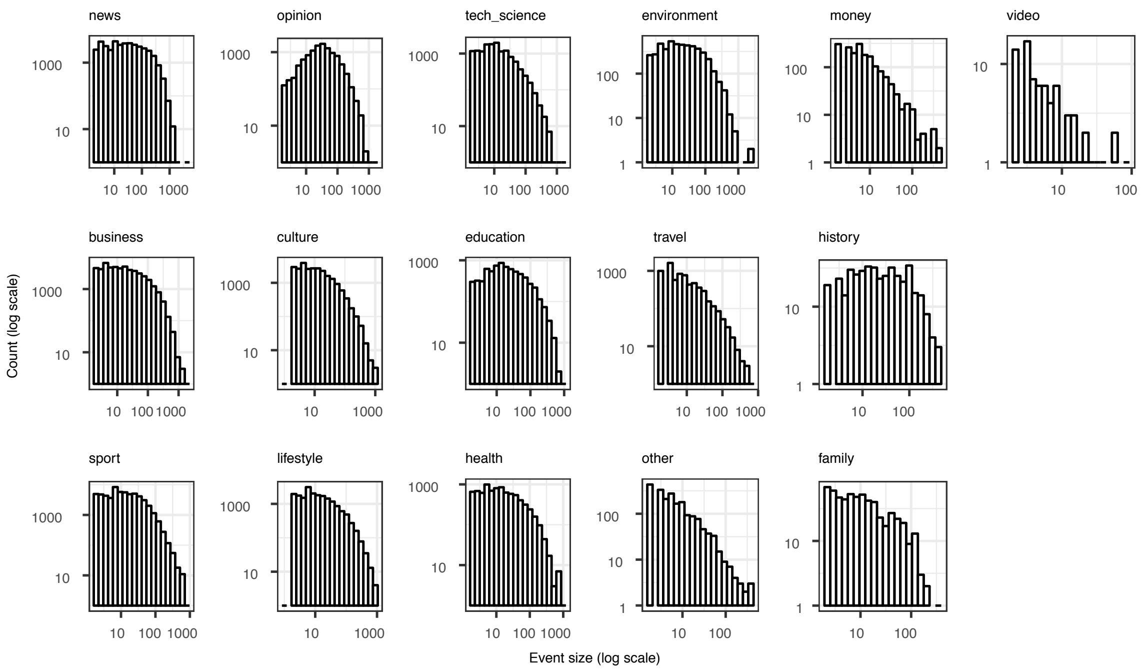


Fig. A.4 Distribution of the event size within the topic categories: Telegraph datataset.

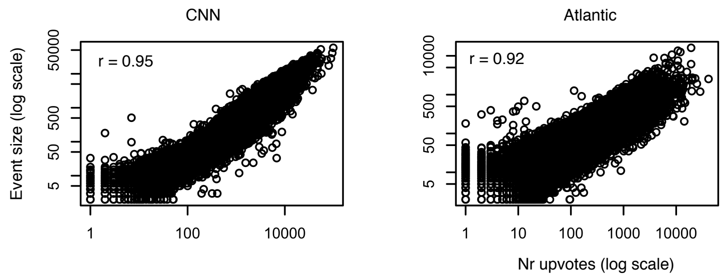


Fig. A.5 Relationship between the event size and the total number of votes in the event. The x axis represent the total number of votes in the event. The y axis represents the event size. Data is pooled from all categories. The Spearman correlation coefficient *r* is computed for the log values. There is a high correlation between the event size and the total number of votes in the event.

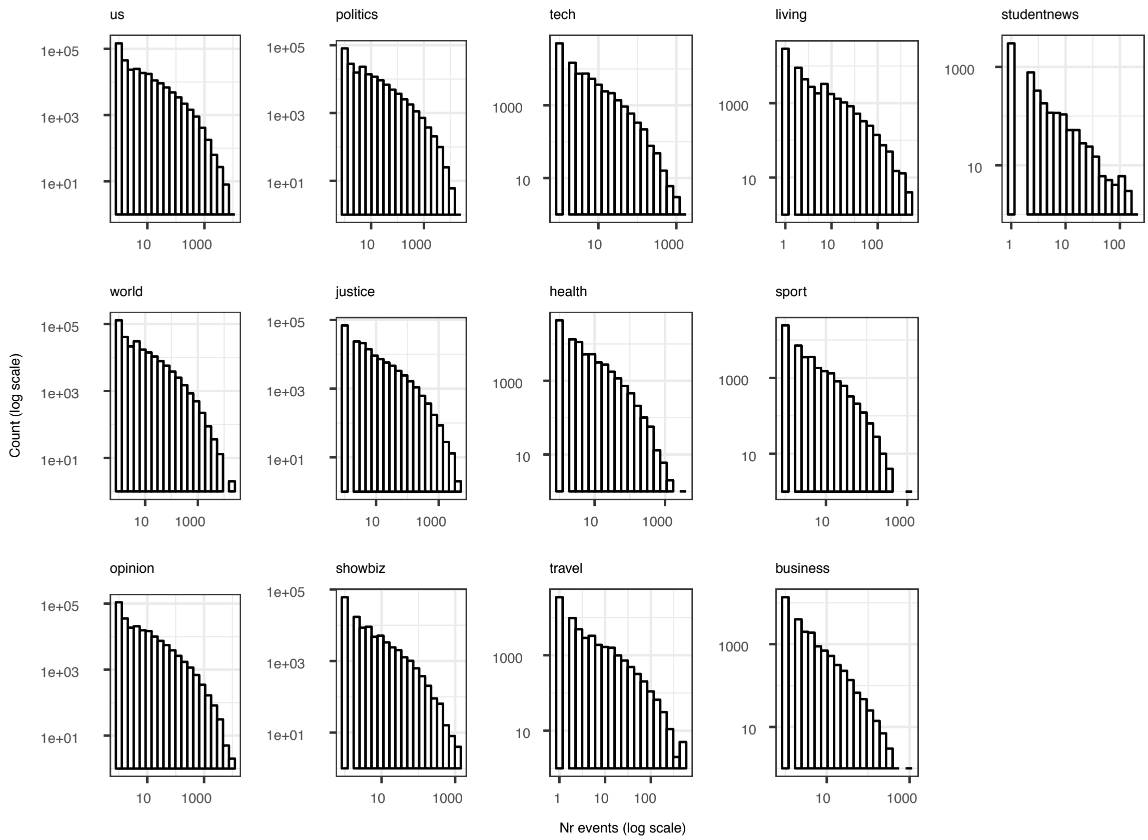


Fig. A.6 Distribution of the number of events in which individuals participated: CNN dataset.

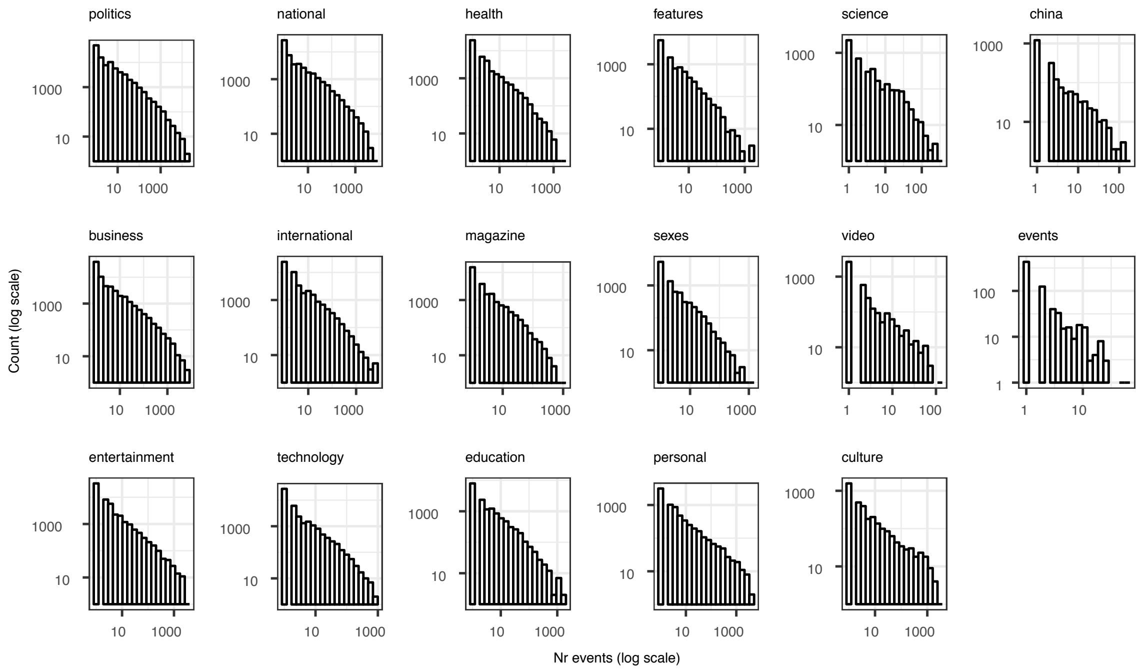


Fig. A.7 Distribution of the number of events in which individuals participated: Atlantic dataset.

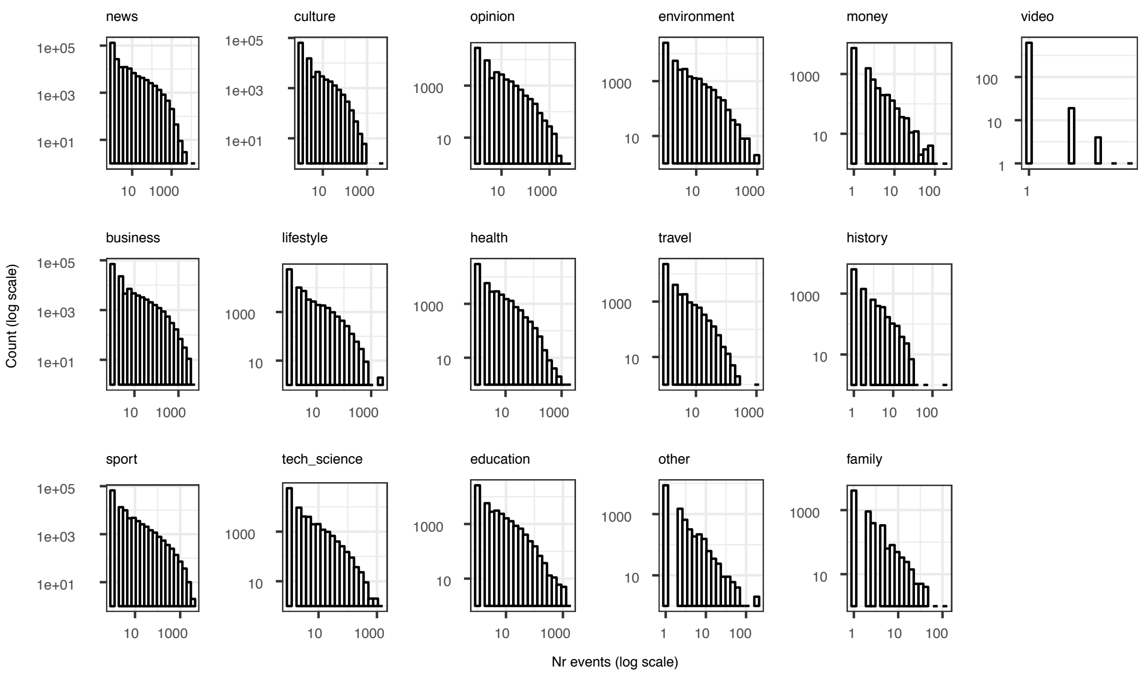


Fig. A.8 Distribution of the number of events in which individuals participated: Telegraph dataset.

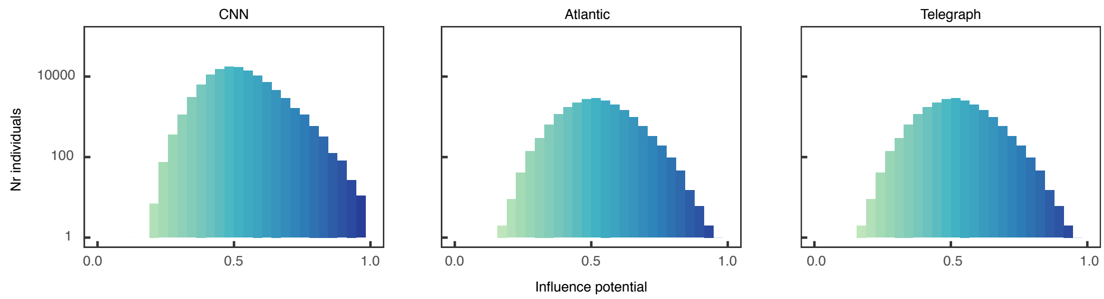


Fig. A.9 **Distribution of the IP under the talkativeness null model**. Data is pooled from all categories. An individual can appear in more than one category. The talkativeness null model leads to the emergence of individuals with high IP.

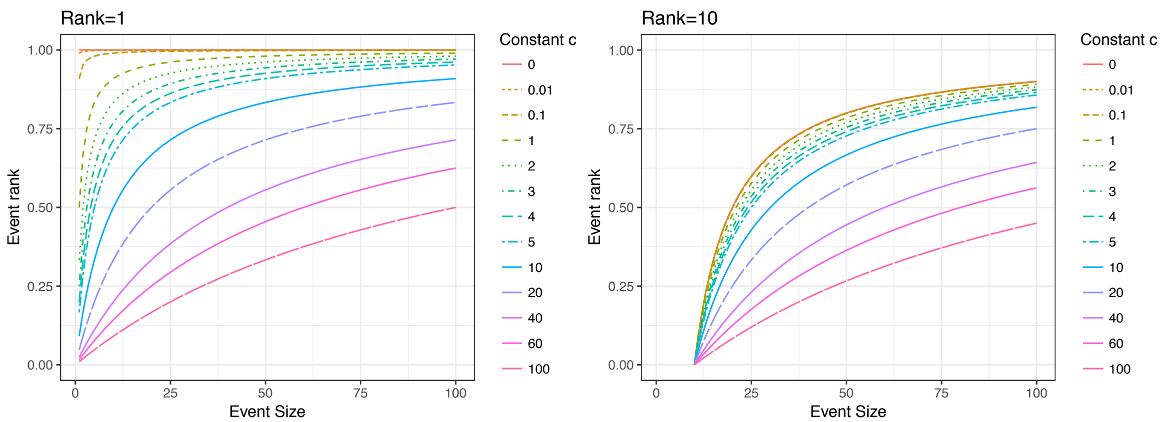


Fig. A.10 **Effect of the constant c on the event rank**. The x axis represents the event size. The y axis represents the event rank for an individual with: the highest number of votes (left panel), 10th highest number of votes (right panel). The lines show the relationship between the event size and the event rank for different values of c.

Appendix B

The effects of social influence and variety seeking on repurchase behavior

B.1 Overview community data

Table B.1 Overview communities

Community	Users	Recipes	Ratings	Categories
I	445,991	39,211	303,881	13
II	203,587	14,486	40,541	14
III	307,727	18,389	119,970	16
IV	150,306	6,044	32,394	16
V	159,766	18,075	68,221	16
VI	115,730	7158	11,349	10
Total	1,383,107	103,373	560,068	-

Category	Recipes	Ratings
Beilagen	862	6564
Desserts	2888	18969
Saucen Dips Brotau	6094	34891
Vorspeisen Salate	1629	15730
Hauptgerichte	4931	59107
Grundrezepte	1382	11492
Backen süß	9487	61317
Suppen	2664	17765
Getränke	2721	17886
Brot Brötchen	3265	31295
Sonstige Hauptgerichte	1584	15303
Backen herzhaft	1356	11703
Baby Beikost Breie	348	1859
Total	39211	303881

Table B.2 Community I. Overview of recipe categories.

Category	Recipes	Ratings
Salse sughi cond	801	2137
Ricette base	232	748
Secondi piatti v	279	1233
Secondi piatti a	1061	3468
Primi piatti	1878	4987
Prodotti da forno	5053	13398
Dessert e pralin	2269	5114
Contorni	403	1515
Bibite liquori e	408	1120
Antipasti	510	1665
Piatti unici	621	1749
Alimentazione in	90	334
Zuppe passati e	401	1566
Pane	480	1507
Total	14486	40541

Table B.3 Community II. Overview of recipe categories.

Category	Recipes	Ratings
Desserts Confis	5065	25895
Sauces dips et p	846	5687
Entrées	1587	9357
Plat principal	1698	14965
Pâtisseries sucr	2554	15599
Viandes	514	5134
Accompagnements	840	7769
Pains Viennoise	970	10702
Tartes et tourte	860	4850
Soupes	1272	6749
Alimentation pou	441	1552
Poissons	416	2920
Plats végétarien	310	2015
Basiques	244	1567
Pâtes Riz	363	3486
Boissons	419	1723
Total	18399	119970

Table B.4 Community III. Overview of recipe categories.

Category	Recipes	Ratings
Verduras y horta	372	1015
Guarniciones y a	109	316
Masas y reposter	685	1994
Dulces y postres	1760	4865
Bebidas y refres	311	469
Carnes y aves	435	1251
Alimentación inf	127	367
Aperitivos y tap	492	1668
Potajes y platos	261	476
Pescados y maris	297	719
Sopas y cremas	383	697
Arroces y pastas	465	1408
Básicas	113	324
Salsas	144	302
Dietas triturada	9	23
Navidad	81	212
Total	6044	16106

Table B.5 Community IV. Overview of recipe categories.

Category	Recipes	Ratings
Baking sweet	4454	16318
Main dishes mea	1733	11958
Soups	999	3563
Sauces dips spr	1708	4704
Basics	686	2575
Pasta rice dish	641	3394
Desserts sweets	3141	8489
Baking savoury	856	3401
Breads rolls	661	3521
Drinks	857	1470
Side dishes	491	2380
Starters	255	847
Main dishes veg	668	2095
Main dishes fis	390	1426
Baby food	124	472
Main dishes oth	411	1608
Total	18075	68221

Table B.6 Community V. Overview of recipe categories.

Category	Recipes	Ratings
Pratos principai	1561	2761
Sobremesas	1680	2487
Bolos e Biscoito	1539	2207
Acompanhamentos	270	431
Prato principal	358	598
Bebidas	383	725
Entradas	359	590
Sopas	391	536
Massas lêvedas	354	578
Crianças	263	436
Total	7158	11349

Table B.7 Community VI. Overview of recipe categories.

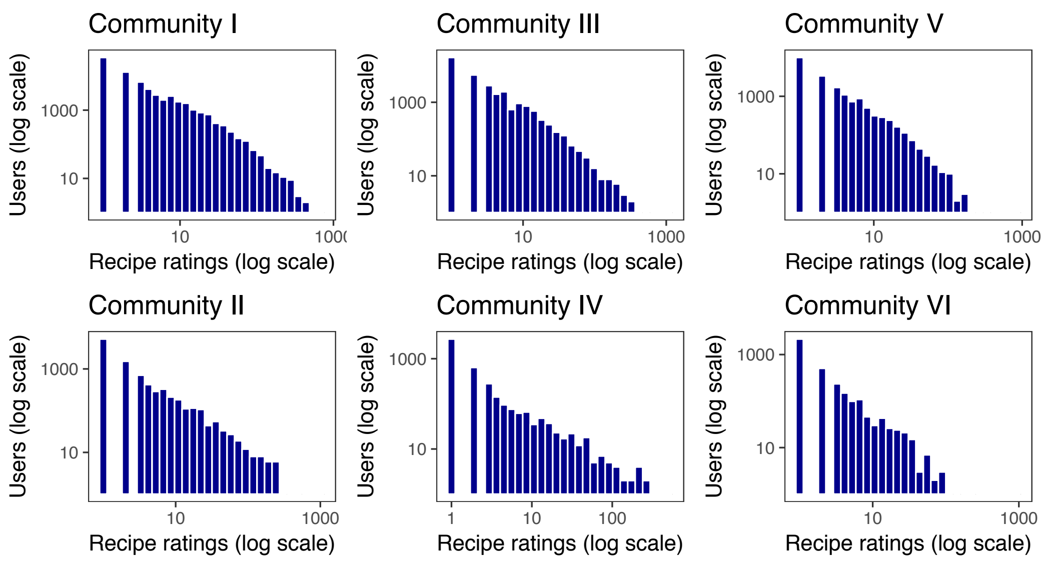


Fig. B.1 Distribution of recipe ratings.

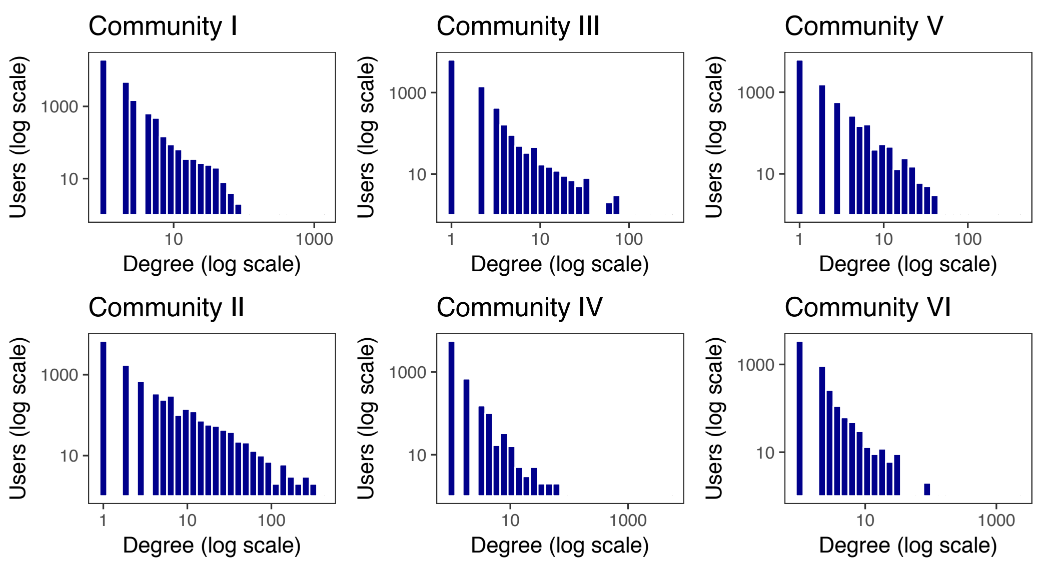


Fig. B.2 Degree distribution.

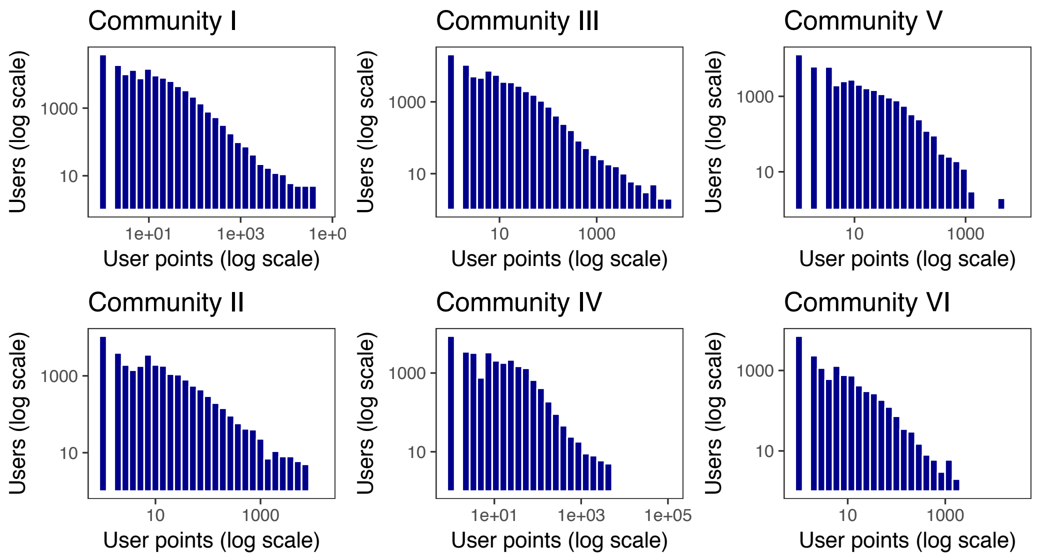


Fig. B.3 Distribution of user points.

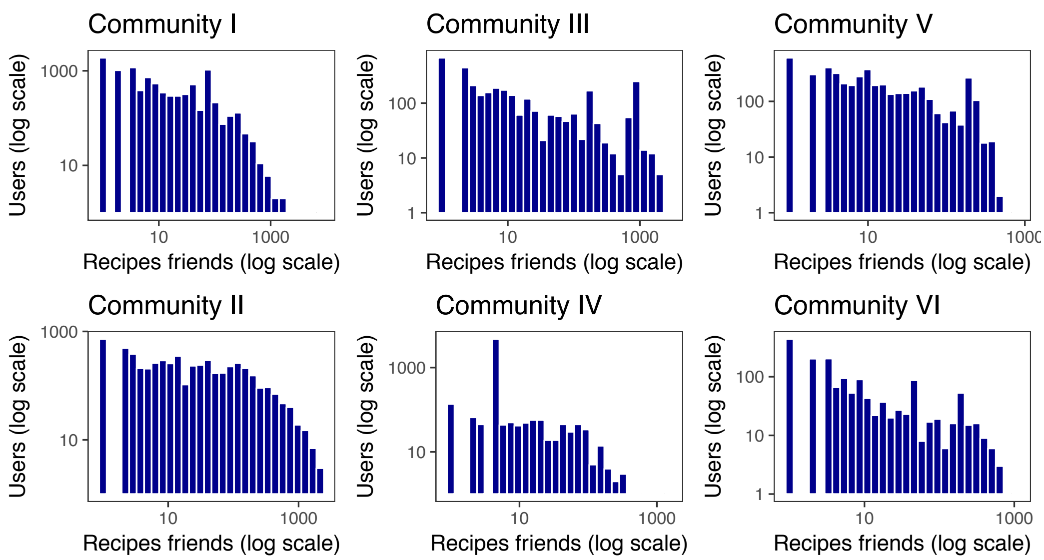


Fig. B.4 Distribution of recipes created by friends.

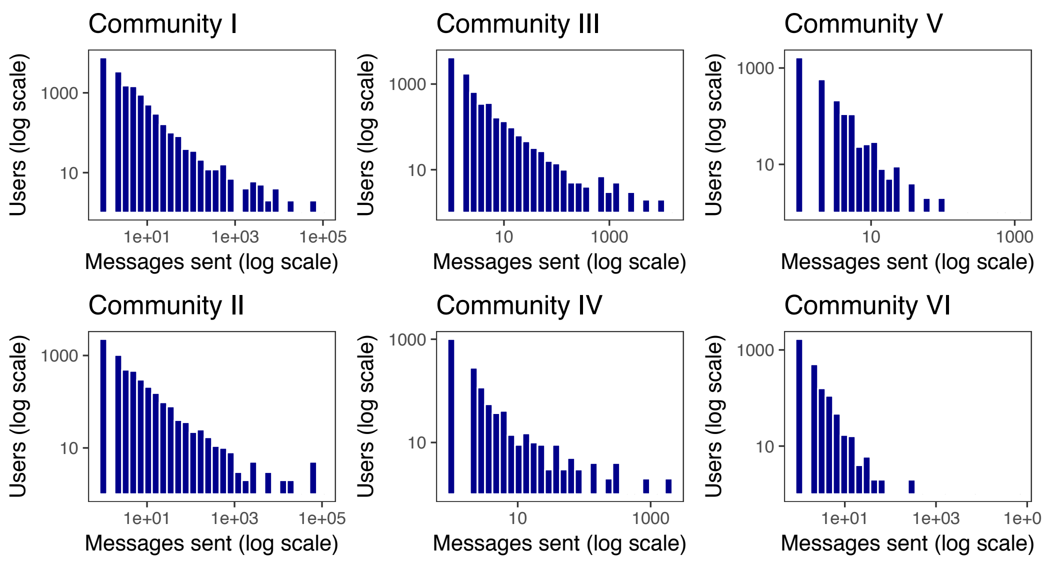


Fig. B.5 Distribution of messages sent.

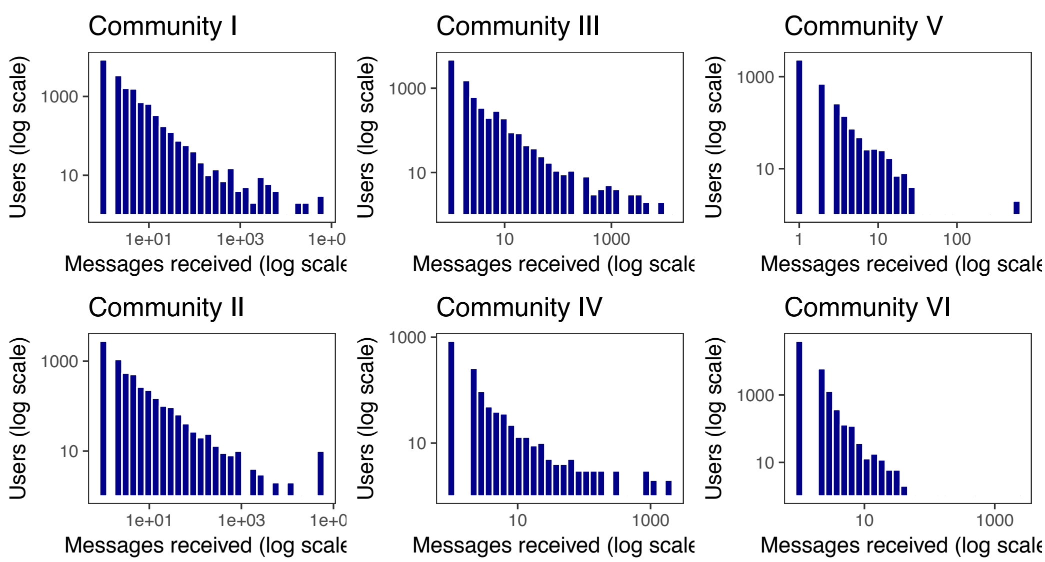


Fig. B.6 Distribution of messages received.

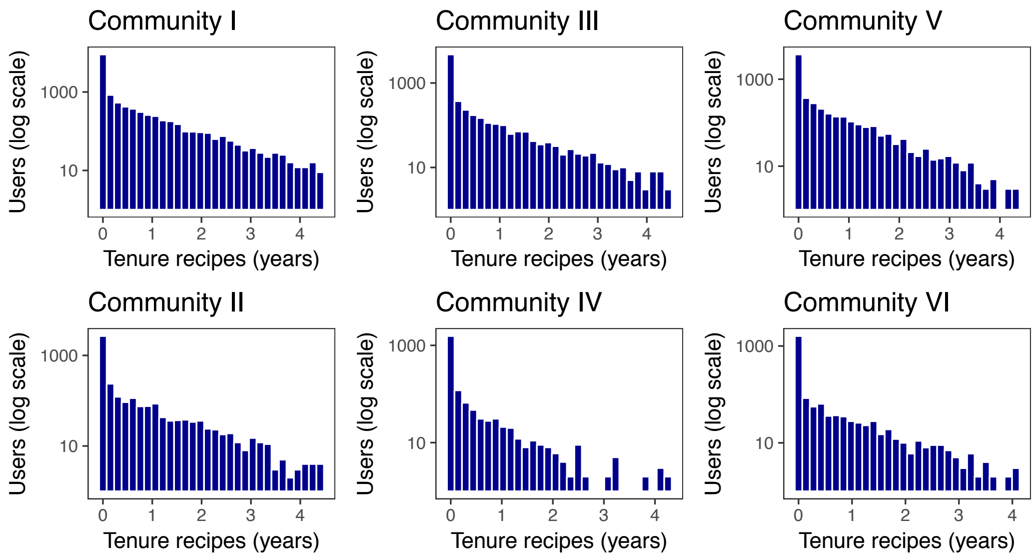


Fig. B.7 Distribution of time during which users create recipes.

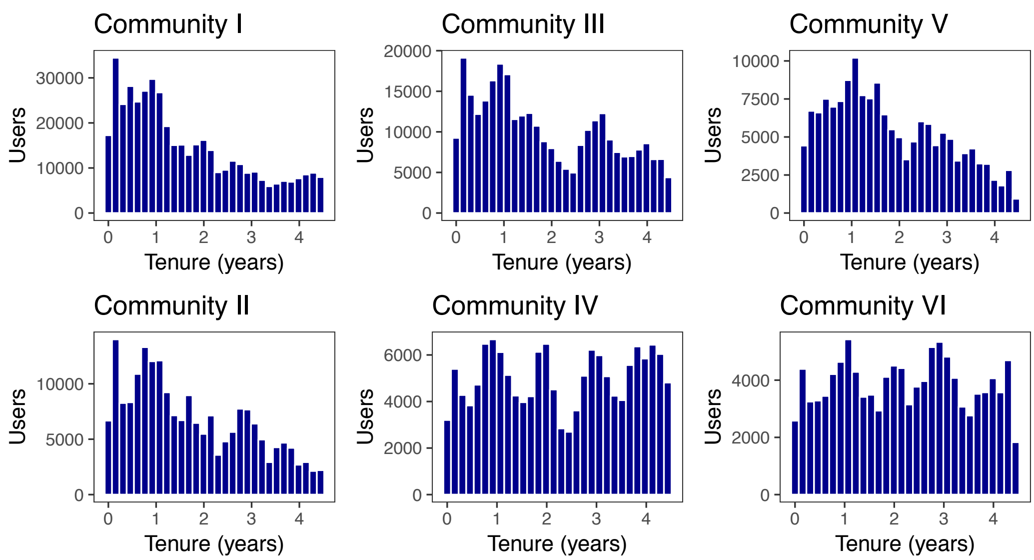


Fig. B.8 Distribution of user lifetime on the platform (in years).

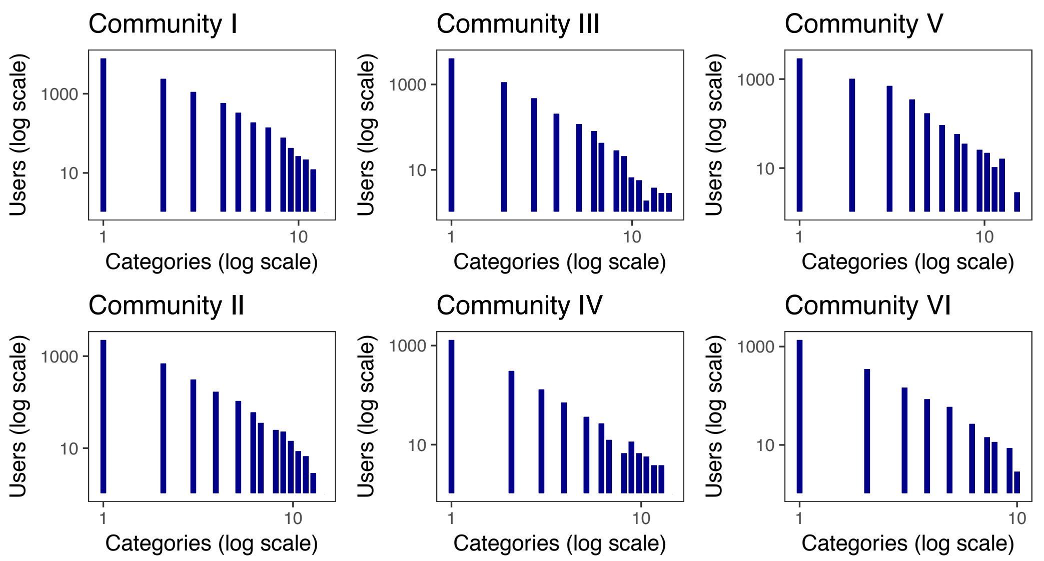


Fig. B.9 Distribution of categories in which users create recipes.

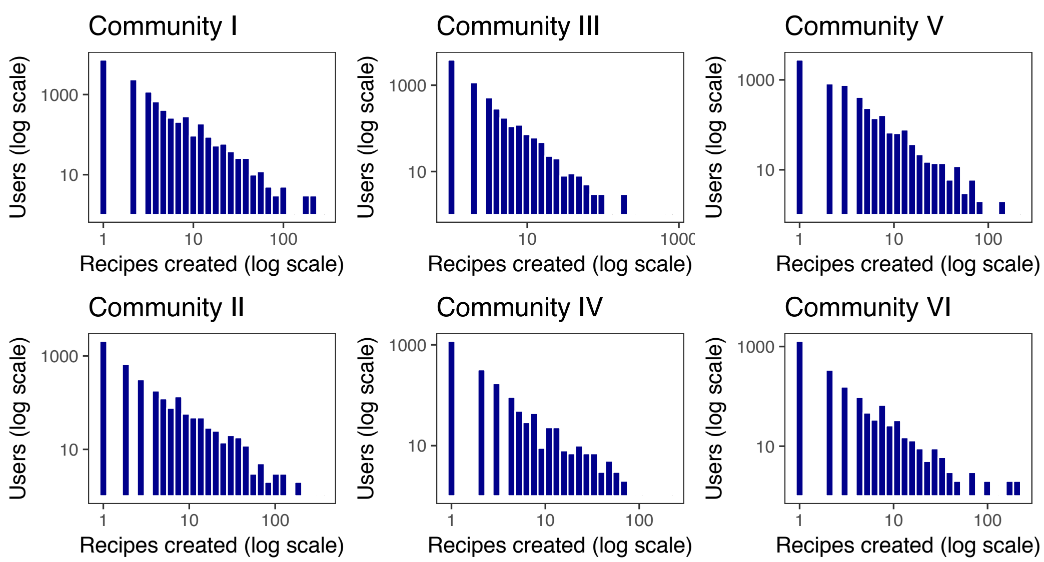


Fig. B.10 Distribution of recipes created.

B.2 Overview aggregated data

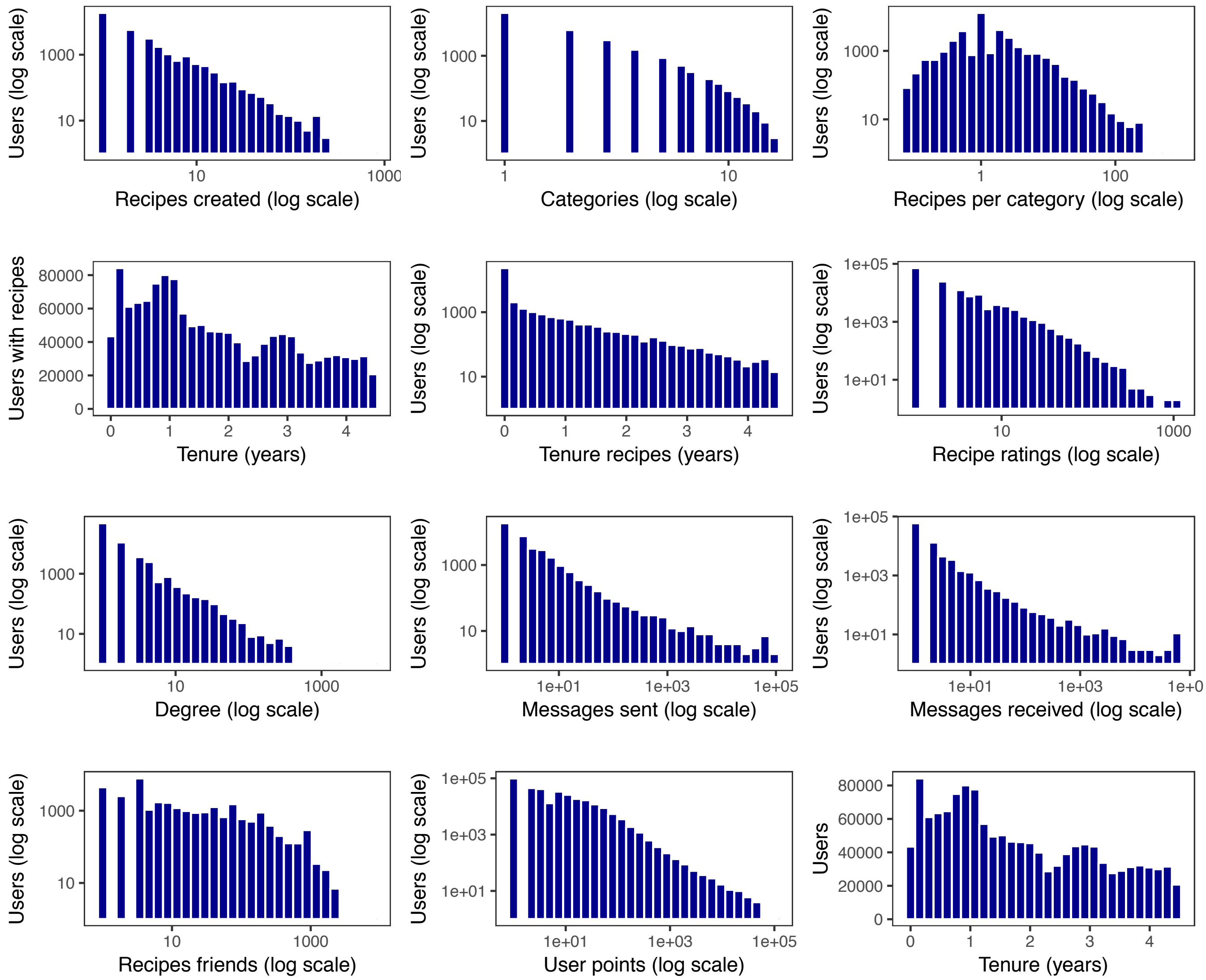


Fig. B.11 Distribution of variables in the aggregated data.

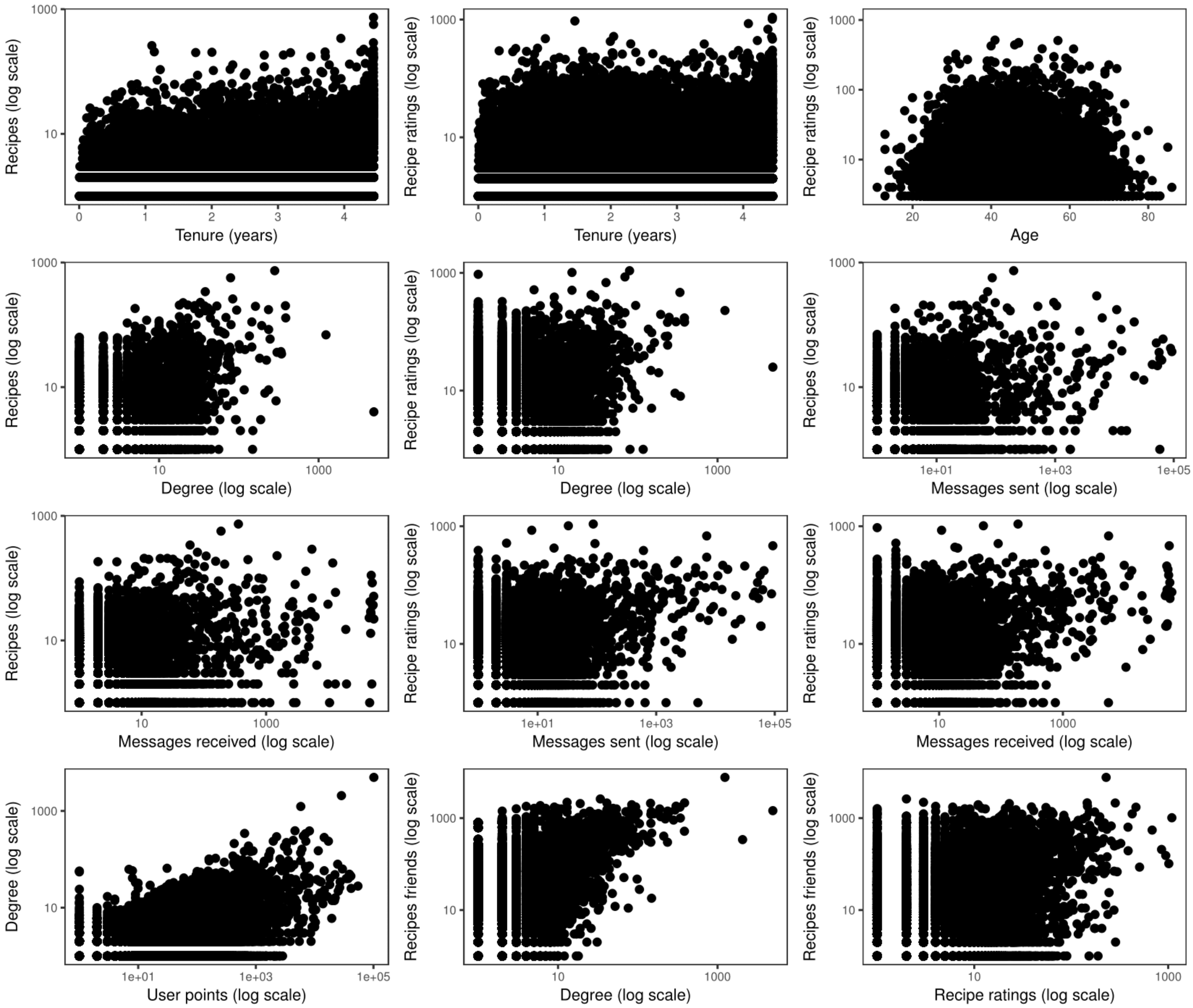


Fig. B.12 **Pairwise relationship between variables in the aggregated data**. For each plot, observations where any of the variables had value zero omitted in order to use logarithmic scales.

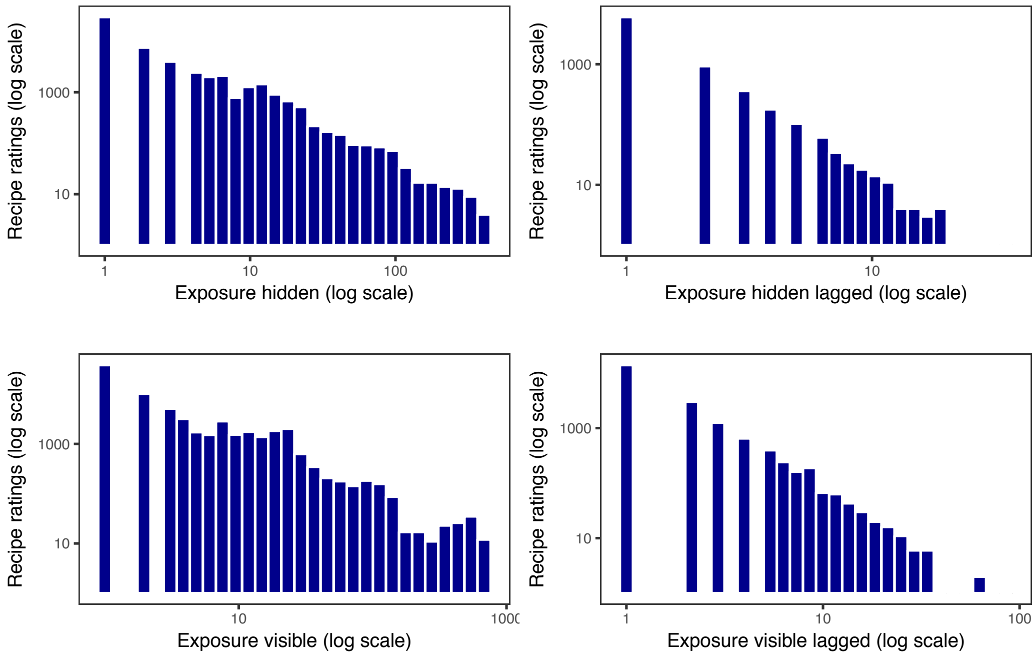


Fig. B.13 Distribution of the exposure variables.

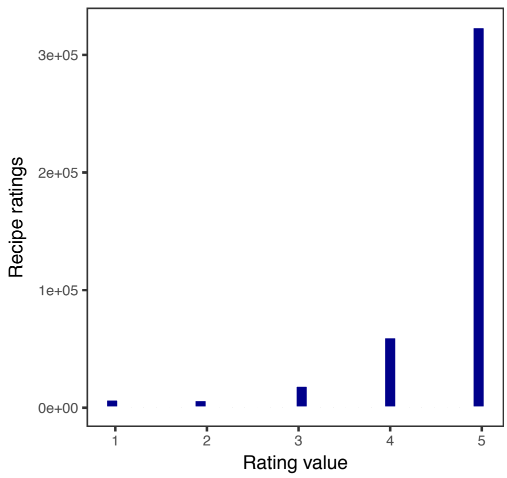


Fig. B.14 **Distribution of the satisfaction variable**. More than 92% of all ratings have a value of at least 4/5.

B.3 Model results

	Coefficient	SE.	p-value
(Intercept)	-1.50	0.06	$< 10^{-10***}$
Exposure hidden	0.34	0.03	$< 10^{-10***}$
Exposure hidden lagged	0.56	0.06	$< 10^{-10***}$
Exposure visible	-0.14	0.03	$< 10^{-5***}$
Exposure visible lagged	-0.44	0.05	$< 10^{-10***}$
Tie strength hidden	0.10	0.06	$0.08^{\cdot}$
Tie strength hidden lagged	0.56	0.12	$< 10^{-5***}$
Tie strength visible	-0.07	0.05	0.17
Tie strength visible lagged	-0.46	0.09	$< 10^{-5***}$
Homophily hidden	1.26	0.57	0.26
Homophily hidden lagged	6.46	0.92	$< 10^{-10***}$
Homophily visible	-0.58	0.55	0.29
Homophily visible lagged	-4.15	0.77	$< 10^{-7***}$
Degree	-0.70 x 10^{-3}	0.39 x 10^{-4}	$0.07^{\cdot}$
Satisfaction	-0.01	0.01	0.22
Category size	0.21 x 10^{-3}	4.08 x 10^{-6}	$< 10^{-10***}$
Available alternatives	-3.11 x 10^{-5}	1.43 x 10^{-6}	$< 10^{-10***}$
Mean inter-event time	-0.98 x 10^{-3}	0.10 x 10^{-3}	$< 10^{-10***}$
Inter-event time	-0.14 x 10^{-3}	6.25 x 10^{-5}	0.02^{*}
Communication volume	-2.09 x 10^{-6}	1.63 x 10^{-6}	0.19
Recipes rated	0.13 x 10^{-3}	0.16 x 10^{-3}	0.41
Recipes created	0.22 x 10^{-2}	0.42 x 10^{-3}	$< 10^{-6***}$
Recent engagement	-0.23 x 10^{-3}	0.18 x 10^{-3}	0.21
Engagement	-4.15 x 10^{-7}	9.98 x 10^{-6}	0.96
Tenure	-0.48 x 10^{-3}	0.01	0.98
Community I	0.09	0.03	0.01^{*}
Community II	0.25	0.04	$< 10^{-7***}$
Community III	-0.09	0.04	0.03^{*}
Community IV	0.54	0.07	$< 10^{-10***}$
Community VI	0.71	0.06	$< 10^{-10***}$

Table B.8 Results model M1.

	Coefficient	SE	p-value
(Intercept)	-1.50	0.06	$< 10^{-10***}$
Exposure hidden	0.24×10^{-1}	0.29×10^{-3}	$< 10^{-10***}$
Exposure hidden lagged	0.04	0.68×10^{-2}	$< 10^{-10***}$
Exposure visible	-0.63×10^{-3}	0.11×10^{-3}	$< 10^{-7***}$
Exposure visible lagged	-0.01	0.25×10^{-2}	$< 10^{-6***}$
Tie strength hidden	0.313	0.05	$< 10^{-7***}$
Tie strength hidden lagged	0.81	0.11	$< 10^{-10***}$
Tie strength visible	-0.17	0.04	$3.19 \times 10^{-4***}$
Tie strength visible lagged	-0.68	0.09	$< 10^{-10***}$
Homophily hidden	2.47	0.56	$1.38 \times 10^{-5***}$
Homophily hidden lagged	6.25	0.92	$< 10^{-10***}$
Homophily visible	-1.17	0.55	0.04^{*}
Degree	-0.96×10^{-3}	0.69×10^{-3}	0.17
Satisfaction	-0.01	0.01	0.29
Category size	0.20×10^{-3}	4.15×10^{-6}	$< 10^{-10***}$
Available alternatives	-3.02×10^{-5}	1.4×10^{-6}	$< 10^{-10***}$
Inter-event time	-0.10×10^{-2}	0.10×10^{-3}	$< 10^{-10***}$
Mean inter-event time	-0.14×10^{-3}	6.23×10^{-5}	0.02^{*}
Communication volume	-3.13×10^{-6}	1.67×10^{-6}	$0.06^{\cdot}$
Recipes rated	0.19×10^{-3}	0.16×10^{-3}	0.23
Recipes created	0.25×10^{-2}	0.43×10^{-3}	$< 10^{-8***}$
Recent engagement	-0.28×10^{-3}	0.19×10^{-3}	0.15
Engagement	-8.22×10^{-8}	1.68×10^{-2}	0.99
Tenure	0.65×10^{-3}	0.01	0.97
Community I	0.10	0.03	0.01^{*}
Community II	0.26	0.04	$< 10^{-8***}$
Community III	-0.08	0.04	$0.05^{\cdot}$
Community IV	0.53	0.07	$< 10^{-10***}$
Community VI	0.72	0.06	$< 10^{-10***}$

Table B.9 Results model M2.

Appendix III

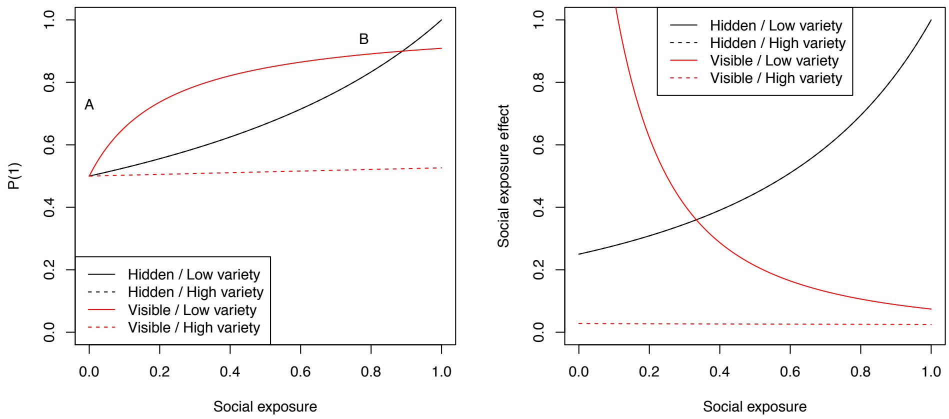


Fig. B.15 **Effect of one social exposure variable on the probability to buy the advocated product when the other has no effect**. The x axis represents social exposure. It can be interpreted as the probability that the second stage decision is to repurchase (hidden) or to switch (visible). **Left panel: The effect of social exposure on the probability to buy the advocated product**. The y axis represents the probability to buy the advocated product. **Right panel: The effect of social exposure on the change in the probability to buy the advocated product**. The y axis represents the social exposure effect (the value of the derivative of P(1) with respect to either E_h or E_v).

B.4 Questionnaires

B.4.1 Study 1: The restaurant choice

Intro Block

Thank you for participating in this study which is funded by Rice University, and conducted in collaboration with researchers from University of Zurich.
Through the brief survey that follows, your answers will be helpful in enhancing our knowledge about how people make decisions.

What will happen in this study?
Your participation in this study will involve completing one survey. We anticipate that your involvement will require about 6-8 minutes. However, you have 30 minutes to complete it, so you do not need to rush.

Are there any potential risks in taking part?
There is no risk in participating in this study. There are no right or wrong answers to any of the questions. Participation in this study is completely voluntary. You are free to decline to participate and to end participation at any time for any reason.

Are there any rewards by taking part?
By taking part and completing the questionnaire, you will earn $0.65. It will be deposited to your Mturk account as soon as we have verified your participation.

What happens to the research data provided?
All of your responses will be confidential and only the researchers involved in this study and those responsible for research oversight will have access to the information you provide.

Will the research be published?
The research results will be written up and published in a peer-reviewed academic journal on a strictly anonymous basis. No confidential data will be included in the data analysis or in the publication of the results.

Who has reviewed this project?
This study is being conducted by R. Tanase & R. Algesheimer from the University of Zurich and U. Dholakia from Rice University. This project has been reviewed by, and received ethics clearance through, the Rice University's Research Ethics Committee. If you have any questions about this study, you may contact U. Dholakia at dholakia@rice.edu.

IMPORTANT NOTICE:
You must be at least 18 years of age to participate in this study.

Do you give your consent to participate in the study under these conditions?

- yes
- no

Energy market description

In this study, we are interested in your decision making about electricity providers. <u>Please read the following description carefully because the rest of the survey depends on your understanding of this context</u>.

In some states in the US, the electricity market is deregulated. Because of this, power generation companies that produce electricity cannot sell electricity directly to consumers.

The electricity companies that sell electricity to end consumers buy it from the power generators and then compete with each other by offering different pricing structures, customer service benefits, and other incentives.

Consumers can choose easily from a list of providers and sign up with them online within minutes. Usually, they have to sign up for a one-year contract.

Task Description

Imagine you have just moved to a state with electricity deregulation and you need to sign up for electricity with one of four companies. All four companies offer a 12-month contract and have the same average price of 8.2 cents per kWh.

Your first task is to choose one of these four electricity companies.

Their descriptions are as follows. Please read the descriptions carefully before answering.

Task I

EDF Energy

EDF Energy is a company headquartered in Houston that provides electricity services to its customers in the deregulated areas of the US. EDF Energy is committed to bringing value to its customers in the form of low prices, renewable energy options and customised online tools.

AB Power

AB Power is an electricity provider offering energy supply in deregulated areas of the US. Headquartered in Houston, the company offers a variety of services for its customers that include: low priced plans, renewable energy and online tools.

Better Energy

Better Energy is a company based in Houston that provides electricity and energy services to customers across the deregulated areas of the US. Better Energy is passionate about creating experiences and solutions tailored to fit your needs, including low priced electricity plans, online tools and renewable energy options.

Airtricity

Airtricity is a Houston based energy provider that offers a variety of electricity products to suit the needs of its consumers. By choosing Airtricity you can benefit from low priced plans, renewable energy options and online tools to manage your plan.

Which electricity company will you choose?

- AB Power
- Better Energy
- EDF Energy
- Airtricity

OL control, Discount I

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, all four electricity companies are ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL control, Discount II

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, all four electricity companies are ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL control, Discount III

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, all four electricity companies are ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL treatment, Discount I

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, EDF Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, AB Power (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Better Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Airtricity (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL treatment, Discount II

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers <u>50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them</u>.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, <u>EDF Energy (the electricity company you picked for the first year)</u> is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer <u>new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them</u>.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, <u>AB Power (the electricity company you picked for the first year)</u> is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer <u>new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them</u>.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, <u>Better Energy (the electricity company you picked for the first year)</u> is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer <u>new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them</u>.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, <u>Airtricity (the electricity company you picked for the first year)</u> is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL treatment, Discount III

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer <u>new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them</u>.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, <u>EDF Energy (the electricity company you picked for the first year)</u> is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, AB Power (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Better Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Airtricity (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Task II

EDF Energy

EDF Energy is a company headquartered in Houston that provides electricity services to its customers in the deregulated areas of the US. EDF Energy is committed to bringing value to its customers in the form of low prices, renewable energy options and customised online tools.

AB Power

AB Power is an electricity provider offering energy supply in deregulated areas of the US. Headquartered in Houston, the company offers a variety of services for its customers that include: low priced plans, renewable energy and online tools.

Better Energy

Better Energy is a company based in Houston that provides electricity and energy services to customers across the deregulated areas of the US. Better Energy is passionate about creating experiences and solutions tailored to fit your needs, including low priced electricity plans, online tools and renewable energy options.

Airtricity

Airtricity is a Houston based energy provider that offers a variety of electricity products to suit the needs of its consumers. By choosing Airtricity you can benefit from low priced plans, renewable energy options and online tools to manage your plan.

Which electricity company will you choose?

EDF Energy
AB Power
Better Energy
Airtricity

EPA Scale

Please indicate your level of agreement with each of the following statements.

I would rather stick with a brand I usually buy than try something I am not very sure of.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

When I go to a restaurant, I feel it is safer to order dishes I am familiar with.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

If I like a brand, I rarely switch from it just to try something different.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

I am very cautious about trying new or different products.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

Even though certain food products are available in a number of different flavors, I tend to buy the same flavor.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

I enjoy taking chances in buying unfamiliar brands just to get some variety in my purchases.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

I think of myself as a brand-loyal consumer.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

When I see a new brand on the shelf, I'm not afraid of giving it a try.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree

I rarely buy brands about which I am uncertain how they will perform.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

I usually eat the same kinds of food on a regular basis.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

Risk

Please indicate how well the following statements describe you.

At work I would prefer a position with a high salary which could be lost easily to a stable position but with a lower salary.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

To achieve something in life one has to take risks.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

If there is a big chance of profit I take even very high risks.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

To gain high profits in business one has to take high risks.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

If there was a big chance to multiply the capital I would invest my money even in the shares of a completely new and uncertain firm.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

I willingly take responsibility in my work-place.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

The skill of reasonable risk taking is one of the most important managerial skills.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

Demographics

Finally, please answer the following three demographic questions:

What is your gender?

- ○ Female
- ○ Male
- ○ Transgender

- Other
- Prefer not to say

What is your approximate age?

- 18-24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 and over
- Prefer not to say

Please indicate your approximate yearly household income before taxes. (Include total income of all adults living in your household.)

- Under $25,000
- $25,001 - $49,999
- $50,000 - $74,999
- $75,000 - $99,999
- $100,000 -$149,999
- $150,000 - $249,999
- $250,000 and over
- Prefer not to say

Thank you for completing the survey. Your completion code is 3178625.

Please click the >> button one more time to submit your response.

B.4.2 Study 2: The electricity market

Intro Block

Thank you for participating in this study which is funded by Rice University, and conducted in collaboration with researchers from University of Zurich.
Through the brief survey that follows, your answers will be helpful in enhancing our knowledge about how people make decisions.

What will happen in this study?
Your participation in this study will involve completing one survey. We anticipate that your involvement will require about 6-8 minutes. However, you have 30 minutes to complete it, so you do not need to rush.

Are there any potential risks in taking part?
There is no risk in participating in this study. There are no right or wrong answers to any of the questions. Participation in this study is completely voluntary. You are free to decline to participate and to end participation at any time for any reason.

Are there any rewards by taking part?
By taking part and completing the questionnaire, you will earn $0.65. It will be deposited to your Mturk account as soon as we have verified your participation.

What happens to the research data provided?
All of your responses will be confidential and only the researchers involved in this study and those responsible for research oversight will have access to the information you provide.

Will the research be published?
The research results will be written up and published in a peer-reviewed academic journal on a strictly anonymous basis. No confidential data will be included in the data analysis or in the publication of the results.

Who has reviewed this project?
This study is being conducted by R. Tanase & R. Algesheimer from the University of Zurich and U. Dholakia from Rice University. This project has been reviewed by, and received ethics clearance through, the Rice University's Research Ethics Committee. If you have any questions about this study, you may contact U. Dholakia at dholakia@rice.edu.

IMPORTANT NOTICE:
You must be at least 18 years of age to participate in this study.

Do you give your consent to participate in the study under these conditions?

- ◯ yes
- ◯ no

Energy market description

In this study, we are interested in your decision making about electricity providers. Please read the following description carefully because the rest of the survey depends on your understanding of this context.

In some states in the US, the electricity market is deregulated. Because of this, power generation companies that produce electricity cannot sell electricity directly to consumers.

The electricity companies that sell electricity to end consumers buy it from the power generators and then compete with each other by offering different pricing structures, customer service benefits, and other incentives.

Consumers can choose easily from a list of providers and sign up with them online within minutes. Usually, they have to sign up for a one-year contract.

Task Description

Imagine you have just moved to a state with electricity deregulation and you need to sign up for electricity with one of four companies. All four companies offer a 12-month contract and have the same average price of 8.2 cents per kWh.

Your first task is to choose one of these four electricity companies.

Their descriptions are as follows. Please read the descriptions carefully before answering.

Task I

EDF Energy

EDF Energy is a company headquartered in Houston that provides electricity services to its customers in the deregulated areas of the US. EDF Energy is committed to bringing value to its customers in the form of low prices, renewable energy options and customised online tools.

AB Power

AB Power is an electricity provider offering energy supply in deregulated areas of the US. Headquartered in Houston, the company offers a variety of services for its customers that include: low priced plans, renewable energy and online tools.

Better Energy

Better Energy is a company based in Houston that provides electricity and energy services to customers across the deregulated areas of the US. Better Energy is passionate about creating experiences and solutions tailored to fit your needs, including low priced electricity plans, online tools and renewable energy options.

Airtricity

Airtricity is a Houston based energy provider that offers a variety of electricity products to suit the needs of its consumers. By choosing Airtricity you can benefit from low priced plans, renewable energy options and online tools to manage your plan.

Which electricity company will you choose?

- ◯ Airtricity
- ◯ EDF Energy
- ◯ Better Energy
- ◯ AB Power

OL control, Discount I

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, all four electricity companies are ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL control, Discount II

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, all four electricity companies are ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL control, Discount III

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, all four electricity companies are ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL treatment, Discount I

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, EDF Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, AB Power (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Better Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 25% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Airtricity (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL treatment, Discount II

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, EDF Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, AB Power (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Better Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers 50% off their first month's electricity bill but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Airtricity (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

OL treatment, Discount III

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, EDF Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, AB Power (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Better Energy (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Now imagine 12 months have passed, during which you were satisfied with the electricity company you picked. You now have the option to either continue the contract with your current company or sign up with a new company. Selecting either option can be conveniently done online within minutes.

All four companies offer new customers their first month's electricity bill free but none of them offer any incentives to current customers who continue their contract with them.

Furthermore, you find out that JD Power & Associates, a leading market research company, has compiled a list of the best retail energy providers in your state. According to their evaluations, Airtricity (the electricity company you picked for the first year) is the only one of the four electricity companies ranked among the top ten best providers in the state.

Here is again the description of the four electricity companies.

Task II

EDF Energy

EDF Energy is a company headquartered in Houston that provides electricity services to its customers in the deregulated areas of the US. EDF Energy is committed to bringing value to its customers in the form of low prices, renewable energy options and customised online tools.

AB Power

AB Power is an electricity provider offering energy supply in deregulated areas of the US. Headquartered in Houston, the company offers a variety of services for its customers that include: low priced plans, renewable energy and online tools.

Better Energy

Better Energy is a company based in Houston that provides electricity and energy services to customers across the deregulated areas of the US. Better Energy is passionate about creating experiences and solutions tailored to fit your needs, including low priced electricity plans, online tools and renewable energy options.

Airtricity

Airtricity is a Houston based energy provider that offers a variety of electricity products to suit the needs of its consumers. By choosing Airtricity you can benefit from low priced plans, renewable energy options and online tools to manage your plan.

Which electricity company will you choose?

EDF Energy
AB Power
Better Energy
Airtricity

EPA Scale

Please indicate your level of agreement with each of the following statements.

I would rather stick with a brand I usually buy than try something I am not very sure of.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

When I go to a restaurant, I feel it is safer to order dishes I am familiar with.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

If I like a brand, I rarely switch from it just to try something different.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

I am very cautious about trying new or different products.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

Even though certain food products are available in a number of different flavors, I tend to buy the same flavor.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

I enjoy taking chances in buying unfamiliar brands just to get some variety in my purchases.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

I think of myself as a brand-loyal consumer.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

When I see a new brand on the shelf, I'm not afraid of giving it a try.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree
○	○	○	○	○

I rarely buy brands about which I am uncertain how they will perform.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

I usually eat the same kinds of food on a regular basis.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
○	○	○	○	○

Risk

Please indicate how well the following statements describe you.

At work I would prefer a position with a high salary which could be lost easily to a stable position but with a lower salary.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

To achieve something in life one has to take risks.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

If there is a big chance of profit I take even very high risks.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

To gain high profits in business one has to take high risks.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

If there was a big chance to multiply the capital I would invest my money even in the shares of a completely new and uncertain firm.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

I willingly take responsibility in my work-place.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

The skill of reasonable risk taking is one of the most important managerial skills.

Does not describe me	Describes me slightly well	Describes me moderately well	Describes me very well	Describes me extremely well
○	○	○	○	○

Demographics

Finally, please answer the following three demographic questions:

What is your gender?

○ Female

○ Male

○ Transgender

Other

Prefer not to say

What is your approximate age?

- 18-24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 and over
- Prefer not to say

Please indicate your approximate yearly household income before taxes. (Include total income of all adults living in your household.)

- Under $25,000
- $25,001 - $49,999
- $50,000 - $74,999
- $75,000 - $99,999
- $100,000 -$149,999
- $150,000 - $249,999
- $250,000 and over
- Prefer not to say

Thank you for completing the survey. Your completion code is 3178625.

Please click the >> button one more time to submit your response.

Appendix C

Controlling complex policy problems

C.1 Results controllability analysis

Variable	Abbreviation	Type	Classification	Cc
birth rate normal	brn	P	critical	0.41
births	b	F	redundant	0.39
births crowding multiplier	bcm	V	intermittent	0.41
births food multiplier	bfm	V	intermittent	0.41
births material multiplier	bmm	V	intermittent	0.41
births pollution multiplier	bpm	V	intermittent	0.41
Capital	Capital	S	redundant	0.41
Capital Agriculture Fraction	Capital Agriculture Fraction	S	redundant	0.38
capital agriculture fraction adjustment time	cafat	P	critical	0.39
capital agriculture fraction indicated	cafi	V	intermittent	0.39
capital agriculture fraction initial	cafini	P	critical	0.39
capital agriculture fraction normal	cafn	P	critical	0.39
capital depreciation	cd	F	redundant	0.42
capital depreciation normal	cdn	P	critical	0.44
capital initial	cini	P	critical	0.42
capital investment	ci	F	redundant	0.42
capital investment from quality ratio	ciqr	V	redundant	0.39
capital investment in agriculture	cia	F	redundant	0.38
capital investment multiplier	cim	V	intermittent	0.44
capital investment rate normal	cirn	P	critical	0.44
capital ratio	cr	V	redundant	0.42
capital ratio agriculture	cra	V	redundant	0.38
crowding	crowding	V	redundant	0.42
death rate normal	drn	P	critical	0.41
deaths	deaths	F	redundant	0.39
deaths crowding multiplier	dcm	V	intermittent	0.41
deaths food multiplier	dfm	V	intermittent	0.41
deaths material multiplier	dmm	V	intermittent	0.41
deaths pollution multiplier	dpm	V	intermittent	0.41
effective capital ratio	ecr	V	redundant	0.38
effective capital ratio normal	ecrn	P	critical	0.39
food coefficient	fc	P	critical	0.39
food crowding multiplier	fcm	V	intermittent	0.39
food per capita normal	fpcn	P	critical	0.39
food per capita potential	fpcp	V	redundant	0.38
food pollution multiplier	fpm	V	intermittent	0.39
food ratio	fr	V	redundant	0.38
land area	la	P	critical	0.44
material standard of living	msl	V	redundant	0.38
nat res matl multiplier	nrmm	V	intermittent	0.44
natural resource extraction multiplier	nrem	V	redundant	0.38
natural resource fraction remaining	nrfr	V	redundant	0.39
natural resource utilization	nru	F	redundant	0.42
natural resource utilization normal	nrun	P	critical	0.44
Natural Resources	Natural Resources	S	redundant	0.41
natural resources initial	nri	P	critical	0.42
Pollution	Pollution	S	redundant	0.38
pollution absorption	pa	F	redundant	0.39
pollution absorption time	pat	V	intermittent	0.41
pollution capital multiplier	pcm	V	redundant	0.41
pollution generation	pg	F	redundant	0.39
pollution initial	poli	P	critical	0.39
pollution per capita normal	pcn	P	critical	0.41
pollution ratio	pr	V	redundant	0.39
pollution standard	ps	P	critical	0.41
Population	Population	S	redundant	0.38
population density normal	pdn	P	critical	0.44
population initial	pi	P	critical	0.39
quality crowding multiplier	qcm	V	intermittent	0.03
quality food multiplier	qfm	V	intermittent	0.41
quality material multiplier	qmm	V	intermittent	0.41
quality of life	ql	V	redundant	0.02
quality of life normal	qln	P	critical	0.03
quality pollution multiplier	qpm	V	intermittent	0.03

Table C.1 Variable list with full names, abbreviations, classification [66], and $c_c(i)$ scores. Variables are ordered alphabetically.

Curriculum vitae

Personal details

Name, Surname:	Tănase, Radu Petru
Date of birth:	24.06.1987

Education

October 2012 - April 2018	Doctoral program in Business Administration University of Zurich, Graduate School of Business, Zurich, Switzerland
September 2010 - September 2012	Master of Science in Statistics ETH Zurich, Department of Mathematics, Zurich, Switzerland
June 2006 - September 2009	Bachelor in Economics Academy of Economic Studies, Bucharest, Romania